



UNIVERSITA' DEGLI STUDI DI MILANO

Facoltà di Medicina e Chirurgia

DOTTORATO DI RICERCA IN  
MEDICINA DEL LAVORO E IGIENE INDUSTRIALE

XXV CICLO

Direttore: Chiar.mo Prof. Giovanni COSTA

# DIFFERENTIAL SUSCEPTIBILITY OF REPETITIVE ELEMENTS TO AIRBORNE POLLUTANTS

Valeria Motta

Relatore: Chiar.mo Prof. P.A. BERTAZZI

1. ABSTRACT	<i>Page 3</i>
2. INTRODUCTION	<i>Page 5</i>
2.1 Human Repetitive Elements	<i>Page 5</i>
2.2 Epigenetic Regulation of Repetitive Elements	<i>Page 10</i>
2.3 Function of Repetitive Elements in the Human Genome	<i>Page 15</i>
2.4 Methylation of Repetitive Elements and Disease	<i>Page 18</i>
2.5 Environmental Exposures Influence on Repetitive Elements	<i>Page 21</i>
3. RESEARCH PROJECT AIMS	<i>Page 25</i>
4. MATERIALS AND METHODS	<i>Page 26</i>
4.1 Study Participants and Exposure Assessment	<i>Page 27</i>
4.2 Sample Preparation	<i>Page 34</i>
4.3 DNA Methylation Analysis	<i>Page 35</i>
4.4 RNA Expression Analysis	<i>Page 40</i>
4.5 Statistical Analysis	<i>Page 43</i>
5. RESULTS	<i>Page 50</i>
5.1 Methylation Analysis from the Three Studies	<i>Page 50</i>
5.2 Methylation Analysis from the Beijing Truck Driver Air Pollution Study (BT DAS)	<i>Page 56</i>
5.3 Expression Analysis from the Beijing Truck Driver Air Pollution Study (BT DAS)	<i>Page 60</i>
6. DISCUSSION	<i>Page 63</i>
7. REFERENCES	<i>Page 70</i>

## 1. ABSTRACT

**Background.** Repetitive elements take up >40% of the human genome and can change distribution through transposition, thus generating subfamilies. Repetitive element DNA methylation has been associated with several diseases and environmental exposures, including exposure to airborne pollutants. No systematic analysis has yet been conducted to examine the effects of exposures across different repetitive element subfamilies.

**Objective.** To evaluate sensitivity of DNA methylation and expression in differentially-evolved LINE, *Alu*, and HERV subfamilies to different types of airborne pollutants.

**Methods.** We sampled a total of 120 male participants from three studies (20 high-, 20 low-exposure in each study) of steel workers exposed to metal-rich particulate matter (measured as PM<sub>10</sub>) (Study 1); gas-station attendants exposed to air benzene (Study 2); and truck drivers exposed to traffic-derived elemental carbon (Study 3). We measured methylation by bisulfite-PCR-pyrosequencing in 10 differentially-evolved repetitive element subfamilies. We evaluate sensitivity of DNA methylation of the same 10 subfamilies and the expression of the most representative and well studied subfamilies *AluSx* and L1HS in a more wide population of 120 individuals (Beijing Truck Driver Air Pollution Study, BTDas) with a well characterized personal exposure levels of PM<sub>2.5</sub> and ambient PM<sub>10</sub>.

**Results.** In the three studies, high-exposure groups exhibited subfamily-specific methylation differences compared to low-exposure groups: L1PA2 showed lower DNA methylation in steel

workers ( $P=0.04$ ) and gas station attendants ( $P=0.03$ ); L1Ta showed lower DNA methylation in steel workers ( $P=0.02$ ); *AluYb8* showed higher DNA methylation in truck drivers ( $P=0.05$ ). Within each study, dose-response analyses showed subfamily-specific correlations of methylation with exposure levels. Interaction models showed that the effects of the exposures on DNA methylation were dependent on the subfamily evolutionary age, with stronger effects on older LINEs from PM<sub>10</sub> (p-interaction=0.003) and benzene (p-interaction=0.04), and on younger *Alus* from PM<sub>10</sub> (p-interaction=0.02).

In the BTDA Study the group analysis showed a significantly lower DNA methylation in the truck drivers group in *AluSx* ( $P=0.02$ ) and MLT1d ( $P=0.01$ ). The dose response analysis confirmed a lower *AluSx* methylation in relation to PM<sub>10</sub> 8-day mean ( $P=0.047$ ) and a lower MLT1d methylation in relation to exposure to PM<sub>2.5</sub> ( $P=0.002$ ), EC ( $P=0.008$ ), ambient PM<sub>10</sub> study day mean ( $P=0.005$ ) and ambient PM<sub>10</sub> 8-day mean ( $P=0.018$ ).

**Conclusions.** The evolutionary age of repetitive element subfamilies determines differential susceptibility of DNA methylation and expression to airborne pollutants.

## **2. INTRODUCTION**

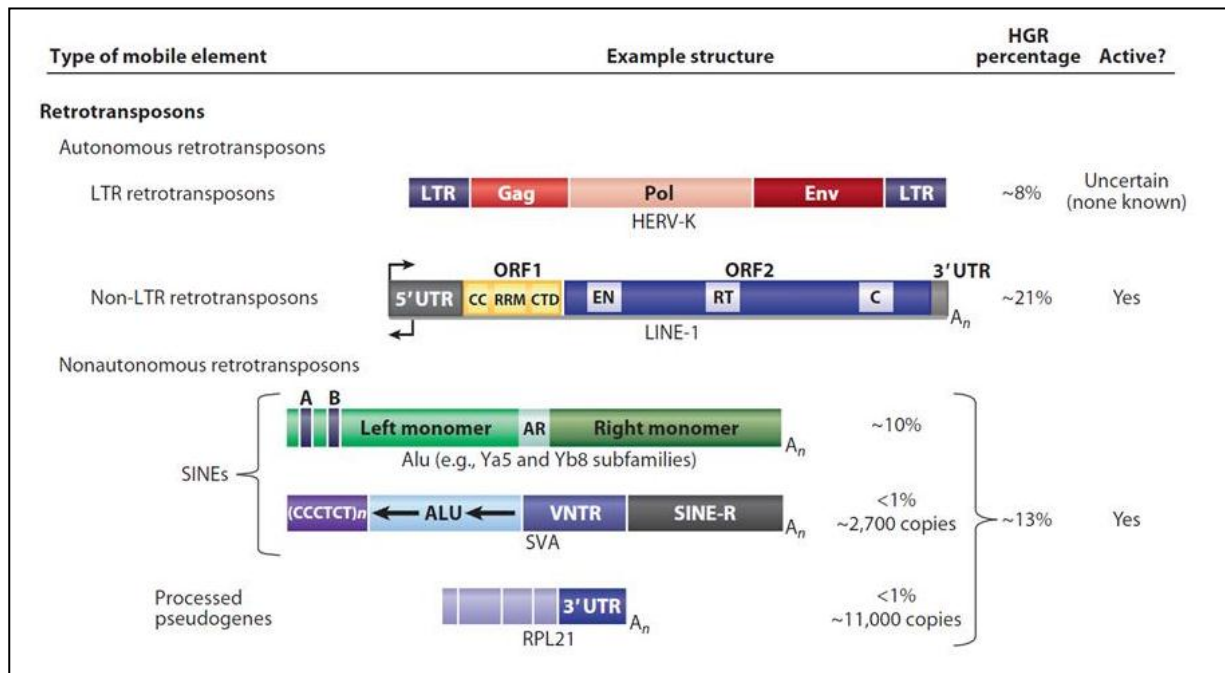
### **2.1 HUMAN REPETITIVE ELEMENTS**

The eukaryotic genome is a complex and dynamic structure. Only about 3% of the mammalian genome is composed of protein-coding sequences compared to about 50% constituted by transposable repetitive elements.

Transposable elements are DNA sequences that are able to jump into new locations within the genome acting as “functional genome reshapers,” which are able to alter gene expression and promote genome evolution [1-3].

In the human genome the most active class elements are termed retrotransposons or retroelements (REs). They move by a “copy-and-paste” mechanism involving reverse transcription of an RNA intermediate and insertion of its cDNA copy at a new site in the host genome. This process is termed retrotransposition. Retrotransposons can be grouped into two major subclasses [4]: retroviral-like or long terminal repeat (LTR) retrotransposons, which include the human endogenous retroviruses (HERVs), and non-LTR retrotransposons, represented in the mammalian genome by long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (**Figure 1**).

**Figure 1.** Transposable element content of the human genome.



### Human Endogenous Retroviruses (HERVs)

HERVs are relics of past rounds of germline infection by exogenous retroviruses that lost their ability to reinfect and became trapped in the genome because they harbor inactivating mutations that render them replication defective. They were found in all vertebrate genomes and constitute about 8% of the human genomic DNA [5]. Most of HERVs sequences have undergone extensive deletions and mutations, therefore becoming transpositionally deficient and transcriptionally silent [6].

### Long Interspersed Nuclear Elements (LINEs)

LINEs are widely distributed in eukaryotes, the LINE-1 (L1) family is covering about 500,000 L1 copies occupying 18% of the haploid genome. L1 elements represent the only family of autonomous non-LTR retrotransposons harboring functional elements that are currently expanding in humans[2]. However, only 80–100 elements are functional and retrotransposition-

competent [7]. A human full-length L1 is 6 kb long and has a 900-nt 5'-untranslated region (UTR) that functions as an RNA polymerase II internal promoter, two open reading frames (ORF1 and ORF2), a short 3'-UTR, and a poly(A) tail. ORF1 encodes a nucleic acid-binding protein that lacks sequence similarity with any other known protein [3, 8]. The ORF2 protein contains endonuclease (EN) and reverse transcriptase (RT) activities as well as a Cys-rich domain, and all three domains are absolutely essential for retrotransposition [9].

L1 retrotransposition is thought to occur by a mechanism termed target primed reverse transcription (TPRT) [10]. During TPRT, L1EN recognizes and cleaves the DNA consensus target sequence 5'-TTTT/AA-3' which means that there are a multitude of potential genomic L1 integration sites [11]. Due to the cis-preference of the L1-encoded protein machinery for its own mRNA, L1 mobilizes preferentially itself. However, in very rare cases, L1s are able to mobilize Alu and SVA RNAs as well as cellular mRNAs whose retrotransposition results in pseudogene formation [12].

#### Short Interspersed Nuclear Elements (SINEs)

SINEs short sequences of 80–500 bp that comprise about 12% of the human genome, and do not code for proteins [13]. SINEs harbor an internal promoter, are pol III-transcribed, and possess at their 3'-end a pA-rich tail. Most SINEs within a given family are full-length and are flanked by TSDs of varying length. Structural similarities between LINEs and SINEs suggested early that the LINE encoded protein machinery is responsible for SINE mobilization. SINEs “hijack” the RT encoded by an autonomous non-LTR retrotransposon for their own mobilization. It is generally accepted that LINEs are used as a source of RT for SINE proliferation [14]. In the human genome, SINEs are represented by two major families termed MIR (mammalian-wide interspersed repeats) and *Alu*. MIR elements are tRNA-like SINEs that include around 470,000

copies constituting about 2% of the human genome; while *Alus* are 7SL RNA-derived elements, include about  $1.1 \times 10^6$  elements occupying 10% of the genome[13]. *Alu* elements are the most abundant repeats in the human genome. The major burst of *Alu* retrotransposition took place 50–60 million years ago (Mya) and has since dropped to a frequency of one new transposition event in every 20–125 births [15].

### *Repetitive elements subfamilies*

Through their “copy and paste” retrotransposition activity, each repetitive element family has gradually accumulated new base substitutions, insertions or deletions referred to as “diagnostic mutations” [16], which are exclusively shared by all subfamily members and allow for distinguishing various subfamilies within each repetitive element family. Through the analysis of these mutations and the divergence of the repetitive elements consensus sequences from the original ancestral sequence, it is possible to estimate an “evolutionary age” for each subfamily [5, 17] (**Figure 2**).

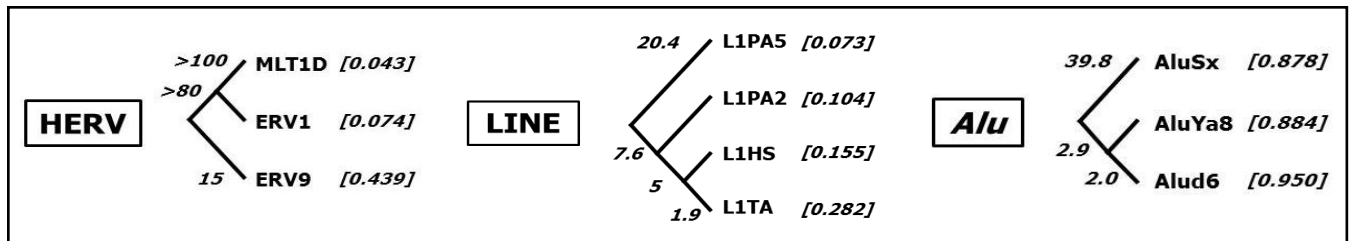
The diagnostic sequence mutations that define subfamilies tend to accumulate hierarchically, so that subfamilies represent an ongoing linear sequential evolution pattern where a series of subfamilies have each been successively derived one from the other. For example, it has been shown that during the past ~40 Mya, all L1 subfamilies in the human genome are derived from a single lineage from which they arose sequentially[18]. Only a few elements called “master elements” are involved in the retrotransposition process and are responsible for the formation of all other subfamily members[19].

The diagnostic mutation rate depends on the CpG density; since CpG di-nucleotides are more sensitive than non-CpG sites they tend to be eliminated through evolution by substitution to



either TpG or CpA. These substitutions are frequent events, as mutation rates in CpG sites is 9.2 fold faster than non-CpG changes [20, 21]. In the mammalian genome, CpG dinucleotides are found to be highly represented in repetitive elements [22]. Through many cycles of substitutions, old subfamilies remain less rich in CpGs and show weakened or no retrotransposon activity, whereas young subfamilies are richer in CpGs and still transcriptionally active in the human genome [21, 23].

**Figure 2.** Cladogram of repetitive elements subfamilies examined. Numbers on the left show the subfamilies age (Mya), numbers in the square buckets represent the CpG density (GpG<sub>o/e</sub>).



## *2.2 EPIGENETIC REGULATION OF REPETITIVE ELEMENTS*

Epigenetic mechanisms are heritable changes not involving an alteration of the genome at the level of nucleotide sequences.

The relationship between repetitive elements and epigenetic regulatory mechanisms starts from the evolutionary link between repetitive elements and their host genomes [24]. Transposition and epigenetics have been intertwined since the discovery of transposable elements. Indeed, the first evidence for epigenetic silencing of transposons arose from experiments that led to their discovery (97). Barbara McClintock found that the sequential breakage and rejoining of chromosomes during maize development resulted in the activation of transposable elements that had been present but quiescent within the maize genome. Because of the disruptive nature of transposition, it is imperative for host genomes to evolve various tools such as DNA methylation to suppress element activity and ensure their own survival. Since the mobilization of repetitive elements has been linked to genomic instability and consequent genetic disorders, mechanisms are thought to have developed in cells [25-27]. This idea forms the core of the “genome defense” model, which proposes that many epigenetic regulatory processes came into existence to defend against the transposition of TEs.

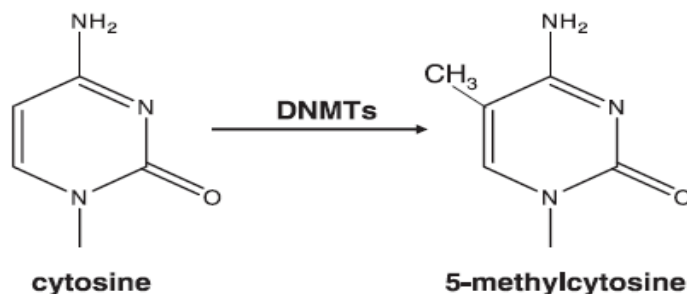
Molecular analysis of these elements revealed that inactivation was associated with DNA methylation, histone modifications and RNA Interference, which provided a link between reversible epigenetic states and molecular modifications [6, 22, 24].

### DNA methylation and DNA methyltransferases

Among the numerous mechanisms of epigenetic silencing in mammals, DNA methylation is the most well studied. DNA methylation represents a stable and reversible covalent modification, heritable by somatic cells after cell division.

5-methyl-cytosine (5MeC) represents 2-5% of all cytosines in mammalian genomes and is found primarily on CpG dinucleotides, often located in enriched regions called CpG islands. Around 60% of the promoters of protein-coding genes in the human genome seem to contain CpG islands, and most of them appear methylated in differentiated tissues [22]. DNA methylation is controlled by DNA methyltransferases (DNMTs), which catalyze the transfer of a methyl group from the methyl donor, S-adenosyl methionine, onto the 5' position of the cytosine ring (**Figure 3**) [28]. They consist of a C terminal catalytic domain and an N terminal regulatory domain. 5 mammalian DNMT's have been identified, however, only DNMT1, DNMT3A and DNMT3B have conclusively been shown to exhibit catalytic activity [29]. DNMT1 is the “maintenance” methyltransferase and its role is to maintain the correct methylation pattern during DNA replication occurring in the cellular cycle. DNMT3a and DNMT3b are responsible for the *de novo* establishment of DNA methylation patterns after implantation [30].

**Figure 3.** Mechanism of DNA methylation.



The extent of DNA methylation changes in an orchestrated way during mammalian development, starting with a wave of demethylation during cleavage (also called the “window of opportunity” step), followed by genome-wide de novo methylation after implantation. The extent of methylation in the gastrulating embryo genome is high owing to the de novo methylation, but it tends to decrease in specific tissues during differentiation [22]. De novo methylation occurs rarely during normal post-gastrulation development but is seen frequently during the establishment of cell lines in vitro and in cancer [31].

The methylation status of a gene is usually inversely correlated with gene expression, so that hypermethylation of certain gene promoters yields gene inactivation, and hypomethylation of these promoters activates or reactivates gene expression [32]. DNA methylation seems to be involved in both initiating gene silencing as well as maintaining it. An increased methylation, in fact, leads to transcriptional repression as it inhibits the binding of transcription factors to their cognate DNA recognition sequences or because it recruits methyl-CpG-binding proteins (MeCPs and MBDs) together with co-repressor molecules [28]. CpG islands in promoters of housekeeping genes are physiologically hypomethylated and therefore transcriptionally active.

#### *Histone remodeling and histone methyltransferases*

In eukaryotic cells, the basic unit of chromatin is a nucleosome consisting of 146 bp of DNA wrapped around two tetramers of core histone proteins H3, H4, H2A and H2B. Histone core and histones tails, especially those of histones H3 and H4, are subject to covalent post-transcriptional modifications, particularly at their lysine and arginine residues. Among others, the most common and well-studied ones are methylation and acetylation, frequently associated with transcriptional

control or localization to specific genomic neighborhoods [33]. Moreover, most of these modifications are cross-regulated. While some histone modifications are mutually exclusive, others promote deposition of different histone marks on other amino acid residues of the same histone tail [34]. Histone H3 tail at lysine 9 (H3K9) can be mono-, diand trimethylated. In animals, H3K9 trimethylation has been implicated in silencing of repetitive elements [35]. At least five H3K9-specific histone methyltransferases (HMTases) that deposit H3K9 methylation marks have been identified in mammals: Suv39 h1, Suv39 h2, Eset/SETDB1, GLP/EuHMTase1 and G9a/EuHMTase2 [36]. However, their niches in regulation of genes and repetitive elements are only beginning to be understood. Histone methylation and deacetylation may also be dictated by DNA methylation, likely promoting further reduction in active and increase in repressive chromatin marks seen on repetitive elements [37, 38].

#### Regulation of repetitive elements by small RNAs

The suppression mechanisms above fall into the category of transcriptional silencing imposed by DNA methylation or histone modifications. Another mechanism of repetitive elements suppression requires mediation by small RNAs. There are several regulatory pathways involving small RNAs that have diverse targets and different levels of regulatory activity. MicroRNAs regulate genes at post-transcriptional and translational levels, while small interfering RNA (siRNA) and Piwi-interacting RNA (piRNA) defend the genome against exogenous and endogenous parasitic elements at transcriptional and post-transcriptional levels.

Regulation of repetitive elements by siRNA RNA interference is a regulatory mechanism initiated by siRNAs through cleavage of dsRNA molecules. The enzyme Dicer, responsible for dsRNAprocessing, generates 21 – 23 nucleotide long siRNA duplexes. siRNA duplexes are

loaded onto the RNA-induced Silencing Complex (RISC). The catalytic component of RISC is an Argonaute protein with a slicer activity directed against the target mRNA molecule bearing a perfect sequence match to the siRNA. mRNAs containing sequences complementary to the original dsRNA are degraded in a process known as PTGS (post-transcriptional gene silencing). The siRNA pathway is conserved in most species where it acts mainly as a defense mechanism against viruses with a dsRNA stage in their lifecycle or against artificially introduced dsRNA. However, siRNA of endogenous origin is a frequent phenomenon in plants and fungi, contributing significantly to transposon silencing [39]. In animals, very recent evidence of such endogenous siRNAs in *Drosophila* (not reviewed here) and in mouse has been reported. The formation of dsRNA requires transcription of both strands. In the case of LINEs and ERVs, this may occur due to presence of active antisense promoters either downstream of the insertion site or contained within the element itself. Small RNAs corresponding to repeats have also been detected in the oocyte, their size of 21 – 23 nt and enrichment in A and U residues at their 5' ends suggesting they are siRNAs [40]. Repeats represented by these siRNAs contain IAP, MT and LINE1 retrotransposon sequences. Recently, a new small RNA pathway, mainly directed against transposons, was described in animals. As their name indicates, piRNAs bind a particular subfamily of Argonautes, the Piwi-like proteins specifically expressed in germ cells, in accordance with the hypothesis that their main role is defense of the genome against transposition [41].

## 2.3 FUNCTION OF REPETITIVE ELEMENTS IN THE HUMAN GENOME

### Repetitive Elements in the Genome Organization

The repetitive element function in the human genome is strictly related to their status and therefore their methylation pattern. Together, the LINE and *Alu* sequences comprise approximately 30% of the genome. Their ubiquitous presence, genomically combined with their relatively conserved sequence and propensity of methyl-CpG targets, highlights their appeal as a representative measure of global methylation [42]. Therefore changes in DNA methylation and chromatin structure at repetitive element sequences carry the potential to influence a variety of cellular functions.

There is evidence that the presence of repetitive elements in the human genome is exploited to organize the genome into active and inactive regions, to separate domains and functional regions within one chromatin domain, to direct transcription and regulate transcript stability, and to respond to cellular stress [43].

Hypomethylation of repeated DNA sequences is expected to lead to the transcriptional activation of those repeat sequences that still contain active promoters. Repeat sequence transcription could potentially impact in a number of ways:

- by disrupting the balance of transcription factors and other regulatory proteins through competitive binding
- through expression of repeat element encoded proteins—transposases, gag, pol or env genes
- through generation of read through transcripts into single copy sequences that encode proteins
- through transcriptional interference by either sense or antisense transcription in relation to neighbouring genes

- through the production of transcripts complementary to endogenous transcripts, generating double stranded RNAs.

For instance many repetitive elements are located in genome regions that are not or only weakly transcribed in normal somatic cells, but many are located within transcriptional units. Theoretically, depending on their orientation, they might disturb transcription [44] by interference (antisense or sense). The presence of LINE sequences in many introns might make functional sense by destabilizing unspliced genomic transcripts [45]. Likewise, the presence of *Alu* elements in the 3'-UTR of many genes may have substantial effects on the stability of their transcripts. The positioning of *Alu* sequences in the 3'-UTR of many genes may even provide a way to regulate mRNA stability through editing of the *Alu* sequences in the transcripts. In some cases where the mechanisms have been more closely investigated, methylation appears to spread from upstream *Alus* into the nearby gene promoter region [43].

#### *Contribution of Repetitive Elements to the Genome Evolution*

The changes arising in the genome can be passed on in the germ-line, unless they are lethal for the germ cell or the individual. In this fashion, retro elements have contributed to human evolution and continue to do so.

It is obvious that the considerable number of repetitive elements inserted in mammalian chromosomes has had a profound impact on the shaping and plasticity of the genomes. Genomic rearrangements caused by scattered homologous proviral sequences gave rise to countless genetic variations on which the evolutionary powers of selection and adaptation could work. Analysis of human genes reveals that mobile elements, including repetitive elements, are overrepresented in the mRNAs of rapidly evolving mammalian genes with a high ratio of non-



synonymous to synonymous mutations, indicative of an increased diversifying selection. Such genes are mostly involved in immunity, stress responses, and responses to external stimuli. These findings point toward an active role for transposable elements in the diversification and expansion of gene families, increasing the speed of evolution in humans and other mammals. Repetitive elements are also useful tools for phylogenetic studies and serve as genomic markers.

### *Repetitive Elements and Disease*

Retrotransposition and recombination involving retroelements in germ cells also contribute to human inherited disease. Some events are passed on within families or in specific populations, while others cause disease in individuals, e.g., cancer resulting from translocations or gene disruption in the germ cell of a parent or during early development. These aspects will be discussed in the section below.

## 2.4 METHYLATION OF REPETITIVE ELEMENTS AND DISEASE

Repetitive elements are largely inactive in normal somatic cells, and most of their activity is restricted to specific phases during the development of germ cells and to the placenta [46] [4]. Hypomethylation in germ cells opens a “window of opportunity” during which transposition of active elements and recombination between active, but also transcriptionally inactive, elements can take place. Retrotransposition and recombination involving repetitive elements in germ cells also contribute to human inherited disease. Some events are passed on within families or in specific populations, while others cause disease in individuals, e.g., cancer resulting from translocations or gene disruption in the germ cell of a parent or during early development. Therefore most retrotransposition was thought to occur in the germline and *De novo* retrotransposition insertions occurring in the germ line are inherited through generations.

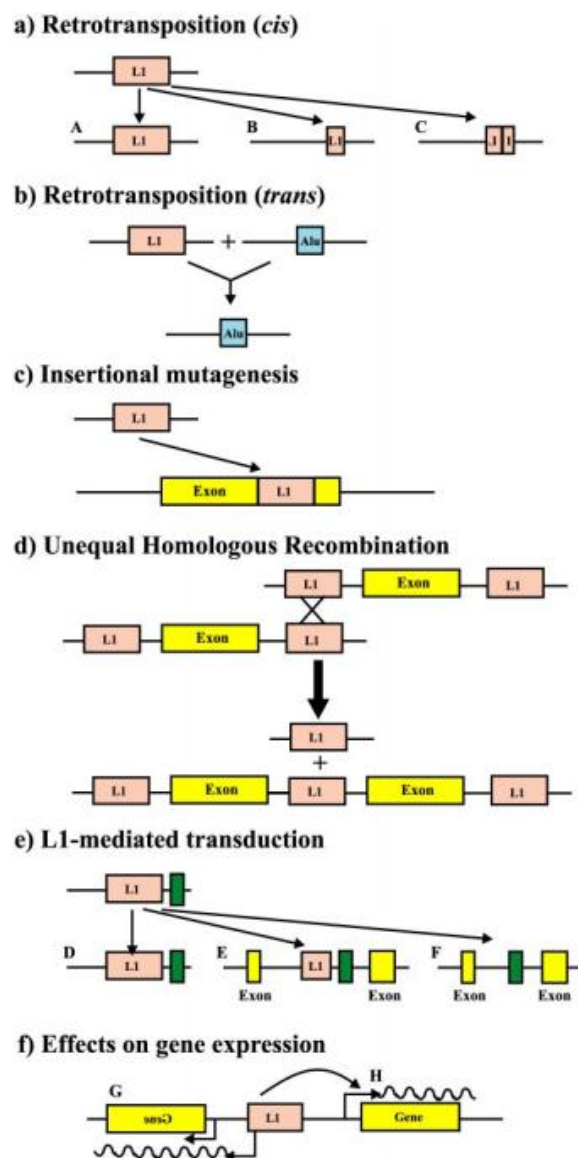
Slowly, over the past several years, evidence has mounted from different experimental systems and from analyses of disease-causing insertions, suggesting that most retrotransposition may be occurring in the soma, and in particular early development. Bulks of newly-inserted repetitive elements have been observed in somatic tissues and, albeit non heritable, have been linked with human disease and cancers [47, 48]. The first report of a somatic L1 insertion was located in the APC gene in an individual with colon cancer [49]. This insertion was present in the cancer and not in the normal colonic tissue. More recently, analyses of a family with a de novo L1 insertion into the CHM gene [50] revealed that the mother of the patient was a germline and somatic mosaic for the L1 insertion causing the disease.

In 2012 Hancks and Kazazian we were able to find 96 retrotransposition events in the literature resulting in single-gene disease [51]. Disease-causing insertions also provide insight into mechanisms by which repetitive elements may alter gene expression. Somatic retrotransposition

may also play a role in tumorigenesis. Nine tumor-specific L1 insertions were identified in 20 lung cancer samples [52].

Retrotransposition can cause human disease by a number of mechanisms, including promoting unequal homologous recombination, direct insertion into genes, and providing the machinery for insertion of other retrotransposons into genes (**Figure 4**).

**Figure 4.** Effects of retrotransposition on mammalian genome.



The majority of the somatic insertions identified in previous studies belong to the known active subfamilies (L1-Ta, AluYa5, AluYb8, and other AluYs) with a subset belonging to the older AluS and L1PA2 subfamilies. AluS and L1PA2 elements are thought to be primarily inactive, because only one disease-causing insertion is associated between the two of them and most L1PA2s do not contain intact ORFs. Data from cell culture has demonstrated that AluS [53] and L1-Ta elements lacking intact ORFs [54, 55] are trans-mobilized by active L1s at a modest level and thus could be retrotransposition-competent sequences *in vivo*.

It has been hypothesized that improper establishment of the epigenetic code can translate to disease development, including loss-of-imprinting disorders and cancer (where this scientific field is most significantly developed) and, more recently, disease-associated states occurring as the result of epigenetic alterations of the germline, termed “epimutations” [56].

In human studies, differences in DNA methylation of LINE-1, *Alu*, and HERV have been consistently demonstrated in response to stress [57] and infections [58], as well as in autoimmune disease [59], cardiovascular disease [60], stroke [60, 61], heart disease [62] and cancers [63].

## 2.5 ENVIRONMENTAL EXPOSURES INFLUENCE ON REPETITIVE ELEMENTS

Recent investigations have repeatedly linked DNA methylation of repetitive elements with environmental exposures. Many groups have evaluated global methylation and air pollution, with most of the studies observing decreased methylation among exposed individuals [64-66], with the exception of one investigation of coke oven workers who had elevated global methylation relative to controls [67]. In human studies, methylation of repetitive elements in blood DNA has been correlated with a number of exposures that generate oxidative stress and inflammation [68, 69], including airborne pollutants [64-66], metals [70-72], and persistent organopollutants [61, 73].

Because of the critical links between genomic hypomethylation and pathogenesis, there is a growing research interest in determining whether changes to the global status of DNA methylation is related to environmental exposures and whether these changes can be biomarkers of disease. In the present work we focused on the analysis of Particulate Matter (PM), Elemental Carbon (EC) and benzene exposure.

### Particulate Matter (PM)

Particulate matter is made up of solid and liquid particles coming from several different sources such as vehicle emissions, industrial and domestic emissions, forest fires, cigarette smoke, natural trees, and climate variations [74].

Besides the different emission sources, PM also differs in size and composition. The size of PM particles is subdivided into several groups depending on their diameter. If the particle diameter is less than 10  $\mu\text{m}$ , it is called PM<sub>10</sub>. PM<sub>10</sub> can be further subdivided into “coarse particles” (PM<sub>10</sub> - PM<sub>2.5</sub>, diameter 10-2.5  $\mu\text{m}$ ), “fine particles” (PM<sub>2.5</sub>, diameter <2.5  $\mu\text{m}$ ), and “ultrafine

particles” (UFPs diameter  $<0.1\ \mu\text{m}$ ). With regards to its composition, PM consists of several metals (Si, Al, Ca, Fe, Ti), transition metals (V, Cr, Ni, Cu, Zn), Pb, inorganic ions ( $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$ ,  $\text{Na}^+$ ,  $\text{NH}_4^+$ ,  $\text{K}^+$ ) and volatile organic compounds coming from industrial chemicals such as fuels, solvents, coatings, feedstocks, and refrigerants. Furthermore, particle composition and concentration are extremely variable and depend on many factors such as climate variations, emission sources, and geographical position [75].

Human previous studies investigated the effects of particulate matter (PM) exposure on global methylation (estimated through *Alu* and LINE-1) in workers of a steel plant with well-characterized exposure to PM with aerodynamic diameters  $<10\ \mu\text{m}$  ( $\text{PM}_{10}$ ). Long-term exposure to  $\text{PM}_{10}$  was negatively associated with methylation in both *Alu* and LINE-1 [66].

### Metals

Several studies have established an association between DNA methylation and environmental metals, including nickel, cadmium, lead, and particularly arsenic [76, 77]. Metal-induced oxidative stress may represent a unifying process to account for these findings across different metals. Metals are known to increase production of reactive oxygen species (ROS) in a catalytic fashion via redox cycling [78]. Oxidative DNA damage can interfere with the ability of methyltransferases to interact with DNA, thus resulting in a generalized altered methylation of cytosine residues at CpG sites [68].

Arsenic is an established carcinogen that lacks carcinogenicity in animal models. Inorganic arsenic is enzymatically methylated for detoxication, using up Sadenosyl-methionine (SAM) in the process. The observation that DNA methyltransferases also require SAM as their methyl donor suggested a role for DNA methylation in arsenic carcinogenesis and other arsenic-related

effects [79]. In rat-liver epithelial cell lines treated with chronic low arsenic doses, Zhao et al. showed malignant transformation associated with depressed SAM levels, global DNA hypomethylation, and decreased DNA methyltransferase activity [79]. Following this findings, several studies have shown that arsenic is associated with global DNA hypomethylation [80-82]. An unexpected finding was recently reported in vivo, as a global dose-dependent hypermethylation of blood DNA was observed in Bangladeshi adults with chronic arsenic exposure. This effect was modified by folate, suggesting that arsenic-induced increases in DNA methylation were dependent from methyl availability [83].

#### Elemental Carbon (EC)

Elemental Carbon is a combustion by-product contained in PM that has been used as a surrogate measure for PM from gasoline- and especially diesel-powered motor vehicles [84]. The terms EC, black carbon (BC) and Carbon Black have been widely interchanged in the literature in the past; EC as the definition of refractory carbon measured by thermal/optical carbon analyzers.

A previous study showed that exposure to black carbon (BC) was also associated with decreased DNA methylation in LINE-1, measured in 1,097 blood DNA samples from the Normative Aging Study (NAS), a repeated measure investigation of elderly men in the Boston area [64].

#### Benzene

Benzene is a widespread airborne pollutant emitted from traffic exhaust fumes and cigarette smoking. Benzene is not toxic and carcinogenic *per se*, but is through its metabolites. Experimental evidences indicate that reactive intermediates are necessary for benzene carcinogenicity and toxicity, but the metabolite(s) responsible are still not fully identified.

Inhaled benzene is partly eliminated in the exhaled air. The remaining is rapidly distributed, crosses blood-brain, placental and gonadal barriers, and is found in several organs including the bone marrow. Benzene metabolites, such as hydroquinone and catechol, reach the bone marrow and can be further resulting in the production of quinones and reactive oxygen species, which bind covalently to biological macromolecules. Un metabolized benzene and several metabolites are eliminated through the kidney and some of them can be measured to assess benzene exposure.

Bollati et al. investigated whether DNA methylation changes are induced by low-benzene exposure in peripheral blood DNA of gasoline station attendants and traffic police officers. High-level exposure to benzene has been associated with increased risk of acute myelogenous leukemia (AML), which is characterized by aberrant global hypomethylation. Airborne benzene exposure was associated with a significant reduction in global methylation measured in LINE-1 and Alu [65]. This findings show that low-level benzene exposure may induce altered DNA methylation reproducing the aberrant epigenetic patterns found in malignant cells.



### **3. RESEARCH PROJECT AIMS**

The primary aim of the study is to evaluate sensitivity of DNA methylation in differentially-evolved LINE, *Alu* and HERV repetitive-element subfamilies to different types of airborne pollutants.

The study will investigate the different susceptibility of repetitive element subfamilies methylation in three matched populations with a well characterized exposure to PM, benzene and EC. In particular the analysis will be done on four LINE subfamilies (L1 PA5, L1 PA2, L1HS, and L1 Ta), three *Alu* subfamilies (*Alu* Sx, *Alu* Yb8, and *Alu* d6), and three HERV subfamilies (MLT1, MLT2, and HERV9).

The secondary aim is to evaluate sensitivity of DNA methylation of the same subfamilies in a more wide population of 120 individuals (Beijing Truck Driver Air Pollution Study, BTDas) with a well characterized personal exposure levels of PM<sub>2.5</sub> and ambient PM<sub>10</sub>.

Among repetitive elements, only LINE-1 and *Alu* have been unequivocally shown to be still active and retrotransposition-competent in the current human genome [19]. Therefore the third aim is to evaluate the expression of the most representative and well studied subfamilies *Alu*Sx and L1HS. Since LINE-1 gene consists of 2 open reading frames coding for proteins involved in the retrotransposition activity, both ORF1 and ORF2 expression will be evaluated.

#### ***4. MATERIAL AND METHODS***

The first part of the study is focused on the methylation analysis of different repetitive element subfamilies. 120 healthy individuals were selected as part of three previous studies in which participants had well-characterized exposure to different airborne pollutants. In Study 1 the participants were recruited from steel workers in Brescia, in Study 2 the participants were selected from a study in Milan and the participants in Study 3 are part of the Beijing Truck Driver Air Pollution Study (BTDAS) in China.

The second part of the study takes in consideration the whole BTDAS participants to look at the repetitive elements methylation in a more wide population and to look at the expression levels of a subset of selected repetitive elements subfamilies.

#### *4.1 STUDY PARTICIPANTS AND EXPOSURE ASSESSMENT*

##### **Study participants and exposure assessment of the three studies**

From each of the three studies, we selected two groups of 20 high-exposed individuals and 20 low-exposed controls matched by age and smoking status. In consideration of the predominance of males in the three studies and to limit potential confounding by gender, we only sampled male participants. **Table 1** shows the characteristics and the exposure levels of the three studies.

In Study 1, the participants were recruited from steel workers in Brescia, Italy [85]. In brief, we selected among workers in a steel production plant 20 individuals with high exposure to metal-rich particles (Particulate Matter with aerodynamic diameter  $>10\ \mu\text{m}$  [ $\text{PM}_{10}$ ]  $\geq 152.2\ \mu\text{g}/\text{m}^3$ ) and 20 individuals with low exposures ( $\text{PM}_{10} \leq 150.0\ \mu\text{g}/\text{m}^3$ ). In the high-exposed group, the mean age of the participants was  $42.4 \pm 7.9$  years, and 12 of them were ex or current smokers. In the low-exposed group, the mean age of the participants was  $37.8 \pm 3.0$  years, and 14 of them were ex or current smokers. PM was measured using a GRIMM 1100 light-scattering dust analyzer (Grimm Technologies, Inc. Douglasville, GA, USA). Measures of airborne PM mass and PM metal components are obtained from 11 work areas of the steel production facility in order to estimate individual exposures.

In Study 2, we selected from a study in Milan, Italy 20 gas-station attendants exposed to airborne benzene as the high-exposed group (air benzene  $\geq 31.2\ \mu\text{g}/\text{m}^3$ ); and 20 office workers as the control group (air benzene  $\leq 23.0\ \mu\text{g}/\text{m}^3$ ) [65]. Seven gas-station attendants and five office workers were ex or current smokers. Mean age was  $39.9 \pm 11.2$  in gas-station attendants and  $39.7 \pm 10.4$  in office workers. The participants wore a passive sampler (stainless steel tube, internal diameter of 9 mm, length of 90 mm) containing Chromosorb 106, near the breathing zone during

the work shift. Air benzene level in the passive sampler was measured by thermal desorption followed by gas chromatography/flame ionization detector analysis.

In Study 3, we selected from a study in Beijing, China 20 truck drivers with exposure of Elemental Carbon (EC), taken as a tracer of traffic particles, as the high-exposed group ( $EC \geq 16.6 \mu\text{g}/\text{m}^3$ ); and 20 office workers as a control group ( $EC \leq 16.1 \mu\text{g}/\text{m}^3$ )[86]. Mean age was  $35.2 \pm 5.1$  years in truck drivers and  $33.4 \pm 6.0$  for indoor office workers. Eight truck drivers and six office workers were ex or current smokers. We measured personal levels of elemental carbon using gravimetric samplers worn near the breathing zone by the study participants during the eight hours of work. Each sampler setup included an Apex pump (Casella Inc., Bedford, UK), a Triplex Sharp-Cut Cyclone (BGI Inc., Waltham, Massachusetts), and a 37-mm Teflon filter placed on top of a drain disc and inside a metal filter holder. The filters were kept under atmosphere-controlled conditions before and after sampling and were weighed with a microbalance (Mettler-Toledo Inc., Columbus, Ohio, USA).

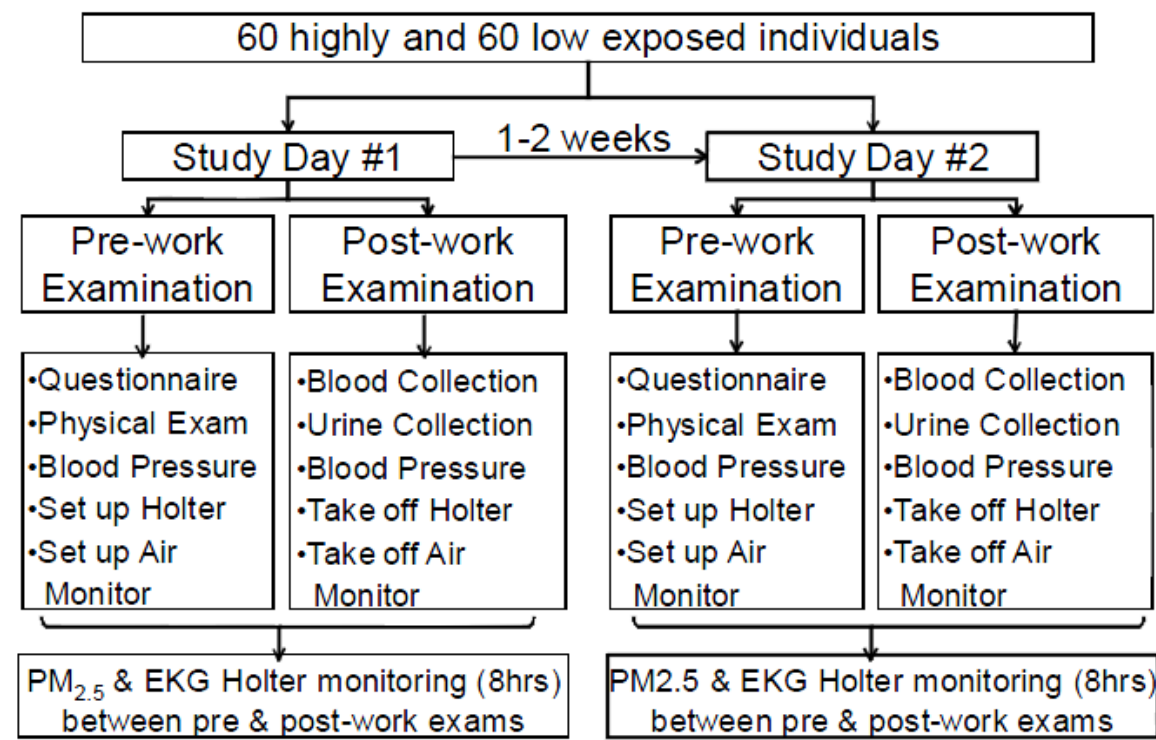
**Table 1.** Characteristics and exposure levels of the three study participants.

		High-exposed steel workers (n=20)	Controls (n=20)
Study 1 Exposure to metal-rich particulate matter (PM)	PM <sub>10</sub> [μg/m <sup>3</sup> ]	Mean ± SD	203.7 ± 22.9
		Range	[152.2 ; 227.9]
	Participants' characteristics	Age [Years], mean ± SD	42.4 ± 7.9
		Ex/current smokers, n (%)	12 (60)
			14 (70)
		Gas-station attendants (n=20)	Controls (n=20)
Study 2 Exposure to air benzene	Air benzene [μg/m <sup>3</sup> ]	Mean ± SD	78.6 ± 42.5
		Range	[31. 2 ; 180.1]
	Participants' characteristics	Age [Years], mean ± SD	39.9 ± 11.2
		Ex/current smokers, n (%)	7 (35)
			5 (25)
		Truck drivers (n=20)	Controls (n=20)
Study 3 Exposure to traffic-derived elemental carbon	Elemental carbon [μg/m <sup>3</sup> ]	Mean ± SD	21.3 ± 4.7
		Range	[16.6 ; 35.6]
	Participants' characteristics	Age [Years], mean ± SD	35.2 ± 5.1
		Ex/current smokers, n (%)	8 (40)
			6 (30)

## Study design and participants of the Beijing Truck Driver Air Pollution Study (BTDAS)

The Beijing Truck Driver Air Pollution Study (BTDAS), conducted shortly before the Beijing Olympic Games (from June 15 to July 27, 2008), included 60 truck drivers and 60 indoor office workers. Because PM levels are highly variable on a day-to-day basis, we examined all subjects on two workdays separated by 1-2 weeks (**Figure 1**). This double sampling scheme allows for detecting short-term variations in cardiovascular and blood measures in relation to the concurrent temporal fluctuations of PM<sub>2.5</sub> levels. Both truck drivers and office workers worked and lived in the Beijing metropolitan area and had been on their current jobs for  $\geq$  two years. The two groups were matched by sex, smoking status and education, and partially matched (5-year intervals) by age (**Table 2**). In-person questionnaire-based interviews were conducted to collect information on demographics, lifestyle, and other exposures.

**Figure 1.** The BTDAS study population and components.



**Table 2.** Characteristic of the BTDas Study Participants.

Characteristics		Office Workers (n = 60)	Truck Drivers (n = 60)	p-value
<b>Sex, n (%)</b>	<i>Male</i>	40(66.67)	40(66.67)	
	<i>Female</i>	20(33.33)	20(33.33)	
<b>Age [Years]</b>		30.27 ± 7.99	33.53 ± 5.67	0.0003
<b>Smoking, n (%)</b>	<i>Never smoker</i>	35 (58.33)	34 (56.67)	
	<i>Ex-smoker</i>	2 (3.33)	2 (3.33)	1.00
	<i>Actual smoker</i>	23 (38.33)	24 (40)	
<b>Cigarette smoked during the study time [cigarettes/day], mean ± SD</b>		2.85 ± 5.21	6.39 ± 9.41	<0.001
<b>BMI [kg/m<sup>2</sup>], mean ± SD</b>		22.76 ± 3.38	24.27 ± 3.21	0.01
<b>Work hour per day, mean ± SD</b>		7.9 ± 1.33	9.88 ± 1.72	<.0001

### Beijing Truck Driver Air Pollution Study (BTDas) exposure assessment

#### Personal exposure measurements

Personal PM<sub>2.5</sub> was measured on both examination days using gravimetric samplers worn by the study subjects during the eight hours of work. The sampler was carried in a belt pack with the inlet clipped near the breathing zone. Each sampler setup included an Apex pump (Casella Inc., Bedford, UK), a Triplex Sharp-Cut Cyclone (BGI Inc., Waltham, Massachusetts), and a 37-mm Teflon filter placed on top of a drain disc and inside a metal filter holder. The filters were kept under atmosphere-controlled conditions before and after sampling and were weighed with a microbalance (Mettler-Toledo Inc., Columbus, Ohio, USA). A time-weighted average of PM<sub>2.5</sub> concentration was recorded by dividing the change in filter weight before and after sampling by the volume of air sampled. The blackness of the same filters used to measure PM<sub>2.5</sub> was assessed

using an EEL Model M43D Smokestain Reflectometer, applying the standard black-smoke index calculations of the absorption coefficients based on reflectance. A factor of 1.0 was assumed for converting the absorption coefficient to EC mass [84, 87], which was then divided by the sampled air volume to calculate average EC exposure concentration. EC is a combustion by-product contained in PM that has been used as a surrogate measure for PM from gasoline- and especially diesel-powered motor vehicles [84] (**Table 3**).

#### Ambient PM<sub>10</sub> data

Ambient PM<sub>10</sub> data during the study period were obtained from the Beijing Municipal Environmental Bureau (<http://www.bjepb.gov.cn/air2008/Air.aspx>). Daily averages of PM<sub>10</sub> computed from data obtained from 27 monitoring stations were used to estimate the average PM<sub>10</sub> level in Beijing. The monitoring stations are distributed across the area to represent Beijing city. Ambient PM<sub>10</sub> data were used to test the hypothesis that the association between particles and blood pressure is with a longer-term average exposure than with the personal monitors. Multiple averaging time windows were used, which included 1-day mean (24 hour average of the day before the examination), as well as 2-day, 5-day, and 8-day means (i.e., average of the 2-8 days before the examination). Daily outdoor temperature data for Beijing city were obtained from the National Oceanic and Atmospheric Administration online database [27] (**Table 3**).



**Table 3.** Level of personal exposure to PM<sub>2.5</sub>, Elemental Carbon (EC) and PM<sub>10</sub> in the BTDA Study.

Exposure	Time window	Office Workers								Truck Drivers								p-value
		N	Mean	SD	10 <sup>th</sup> pct	25 <sup>th</sup> pct	Median	75 <sup>th</sup> pct	90 <sup>th</sup> pct	N	Mean	SD	10 <sup>th</sup> pct	25 <sup>th</sup> pct	Median	75 <sup>th</sup> pct	90 <sup>th</sup> pct	
Personal PM <sub>2.5</sub> <sup>1</sup> (µg/m <sup>3</sup> )	8 hours	120	94.6	64.9	22.4	48.5	86.2	126.6	183.4	119	126.8	68.8	46.3	73.9	116.8	160.5	213.9	<0.001
Personal EC <sup>1</sup> (µg/m <sup>3</sup> )	8 hours	118	13	4	7.3	10.1	13.2	15.7	18.4	120	17.2	6.6	9.2	12.9	16.7	20.8	25.5	<0.001
Ambient PM <sub>10</sub> (µg/m <sup>3</sup> )	Study day mean <sup>2</sup>	120	116.7	50.2	60	82	114	141	181	120	123.5	50.1	72	88	116	150	188	0.29
	8-day mean <sup>3</sup>	120	119.5	23	84.9	101.8	119.9	141.5	146.5	120	120.2	21.5	95.6	102.8	120.4	139	146.3	0.81
Outdoor relative humidity (%)	Study day mean <sup>2</sup>	120	20.6	2.1	17.5	20	21	21.5	23	120	20.6	2.1	17	20	21	22	23	0.93
	8-day mean <sup>3</sup>	120	19.9	2	16	18.8	20.6	21.4	21.7	120	19.8	2	16	18.3	20.5	21.4	21.9	0.81
Outdoor temperature (°C)	Study day mean <sup>2</sup>	120	25.4	2.5	22	23	26	28	29	120	25.3	2.5	22	23	26	28	28	0.96
	8-day mean <sup>3</sup>	120	25	1.4	23.1	24.1	24.6	26.4	27	120	24.9	1.4	23.1	23.6	24.5	26.3	26.9	0.34

<sup>1</sup>Measured during the 8 work hours of examination days using light- weight personal monitor.

<sup>2</sup>24 hours average measure of ambient PM<sub>10</sub>/ relative humidity/temperature on the day of the examination.

<sup>3</sup>24 hours moving average measure of ambient PM<sub>10</sub>/ relative humidity/temperature on the 8 days before the day of examination.

## *4.2 SAMPLE PREPARATION*

Sample preparation was conducted in all studies using standardized procedures. Buffy coat from whole blood was collected from each participant and instantly stored at -80°C until genomic DNA was isolated. Blood for RNA isolation was collected only from the BTDas participants; 2.5 ml of blood were sampled in a PAXgene Blood RNA tubes (PreAnalytiX) and stored at -80°C until RNA was isolated.

### **DNA extraction**

Genomic DNA from all participants in the study was isolated with the same batch of reagents in narrow time windows in order to minimize technical and operator variations. DNA was extracted using the Wizard Genomic DNA purification kit (Promega, Madison, WI) following the manufacturer's instructions. DNA concentration was quantified using ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Pure DNA should have a ratio of OD<sub>260</sub>/OD<sub>280</sub> of approximately 1.8. A ratio that is very different from 1.8 (i.e., <1.5 or >2) may be indicative of either residual protein or organic solvents in the DNA sample. DNA integrity was verified by gel electrophoresis. DNA integrity was verified by gel electrophoresis.

### **RNA extraction**

RNA was extracted using the PAXgene Blood RNA Kit (PreAnalytiX, Quagen, Valencia, CA ) with DNase treatment according to the manufacturer's instructions. RNA concentration was quantified using ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). An Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) was used to assess RNA integrity based on the RIN (RNA Integrity Number) factor; presence of low molecular weight RNA (5S) was also verified.

### 4.3 DNA METHYLATION ANALYSIS

We examined DNA methylation of four LINE-1 subfamilies (L1PA5, L1PA2, L1Hs, and L1Ta) [88], three *Alu* subfamilies (*AluSx*, *AluYb8*, and *AluYd6*) [89], and three HERV subfamilies (MLT1D, ERV1, and ERV9) selected within each family to represent different evolutionary ages ranging from old to young elements [90].

The repetitive element evolutionary age is defined as the estimated time at which each element inserted in the human genome. We identified an evolutionary age (expressed in Mya [million years ago]) for each repetitive element subfamily based on information available in previous literature [18, 88-91]. As a measure of CpG density we calculated the ratio between observed and expected CpG content (CpG[o/e]) for each subfamily. The CpG observed/expected (CpGo/e) ratio is calculated by formula;  $((\text{Num of CpG}/(\text{Num of C} \times \text{Num of G})) \times \text{Total number of nucleotides in the sequence})$  [92] (**Table 4**).

**Table 4.** Characteristics of the repetitive elements subfamilies examined in the present study

	Repetitive Element	Evolutionary Age (Mya)	CpG density (CpG <sub>o/e</sub> )
LINE-1	L1PA5	20.4	0.073
	L1PA2	7.6	0.104
	L1Hs	5	0.155
	L1Ta	1.9	0.282
<i>Alu</i>	<i>AluSx</i>	40	0.878
	<i>AluYb8</i>	2.9	0.884
	<i>AluYd6</i>	2	0.95
HERV	MLT1D	98.2	0.043
	ERV1	24.4	0.074
	ERV9	15	0.439

To study the genome global methylation, Yang proposes a method based on the DNA treatment with bisulfite and subsequent PCR amplification. Repetitive elements are usually heavily methylated and so they can be used as a marker for genome global DNA methylation.

A different forward primer is used, it is labeled with biotin and so allows the separation of the final PCR product on sepharose beads. Biotinylated PCR products can be purified and denatured to single strand DNA to be used as a template in the pyrosequencing reaction.

### **Bisulfite treatment**

In order to be able to maintain methylation pattern through PCR amplification, DNA has to be treated with sodium bisulfite. Sodium bisulfite can selectively deaminate cytosine but not 5-methylcytosine to uracil, causing a stable modification in DNA. So, cytosines that were originally methylated will be amplified as cytosines, while unmethylated cytosines will be transformed to uracils and then be amplified as thymines.

1 µg of genomic DNA was treated using the EZ DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer's protocol. Final elution was performed with 30 µL M-Elution Buffer. Bisulfite-treated DNA was aliquoted and stored at -80°C until ready for use.

### **PCR and pyrosequencing**

The PCR reaction takes advantage of a forward primer labeled with biotin in order to allow the separation of the final PCR product on sepharose beads. Biotinylated PCR products can be purified and denatured to single strand DNA to be used as a template in the pyrosequencing reaction.

Pyrosequencing is an highly quantitative DNA sequencing method developed by M. and P. Nyren Ronago in 1996. It's based on the sequencing of a single DNA chain by the synthesis of

the complementary strand; every time a nucleotide is incorporated into the growing fragment a cascade of enzymatic reactions is activated and this lead to the production a light signal. The visible light generated is proportional to the number of incorporated nucleotides. Although a wide variety of different methods are available, pyrosequencing is the only one that permits an accurate quantification for every single C site and a very intra-run low variability, which is necessary to study healthy tissue exposed to environmental pollutants [93].

The PCR and Pyrosequencing primer sequences for L1HS, *AluSx*, and *AluYb8* have been previously published by Yang *et al.* [42] and by Choi *et al.* [94]. We developed additional assays specific for three LINE-1 subfamilies (L1PA5, L1PA2, and L1Ta), for *AluYd6* subfamily, and for three HERV subfamilies (MLT1D, ERV1, and ERV9) (primer sequences and PCR conditions are listed in **Table 5**). A minimum number of 1 CpG to a maximum of 5 CpGs were evaluated in each assays. For the L1HS element a target region inside the CpG-rich region of the 50 internal promoter including 5 CpGs was chosen, whereas for the *AluSx* element we selected a target CpG-rich region near its 30 end encompassing 4 CpG sites. The primers for the other subfamilies were designed in the consensus sequence in order to amplify a global pool of repetitive elements for each subfamily rather than a single element or genomic locus. PCR amplification was performed at standard conditions using the GoTaq® Hot Start Polymerase (Promega, Madison, WI).

**Table 5.** Primer sequences and PCR conditions for methylation analysis.

Family	Assay	Forward Primer (5' to 3')	Reverse Primer (3' to 5')	Sequencing Primer (5' to 3')	Annealing Temperature (°C)
<b>LINE-1</b>	L1PA5	TTAGTTAAGGGAAGAGGGGATAAA	Biotin)ATAAACATAAAACCCT CTAAACCAAACA	TTAGTTAAGGGAAGA	40
	L1PA2	TTAGATAGTGGGYGTAGGTTAGTGGG T	Biotin)CCTCCRAACCAAATAT AAAATATAATCT	GAGTTAAAGAAAGGG	55
	L1Hs	TTTTGAGTTAGGTGTGGGATATA	Biotin)AAAATCAAAAAATTCC CTTTC	AGTTAGGTGTGGGATA TAGT	56.3
	L1Ta	GGGTTAGGGAGTTTTTTTTT	Biotin)CTCTAAACCAAATATA AAATATA	GGGTTAGGGAGTTTTT TTTT	55
<b>Alu</b>	<i>AluSx</i>	Biotin)TTTTTATTAAAAATATAAAAT T	CCCAAACATAAAATACAATAA	AATAACTAAAATTACA AAC	50
	<i>AluYb8</i>	Biotin)AGATTATTTTGGTTAATAAG	AACTACYAACTACAATAAC	AATAACTAAAACACTACA AAC	53.9
	<i>AluYd6</i>	Biotin)GAGATTAYGGTGAAATTT	CCCAAACAAAAATACTATAA	AATAACTAAAACACTACA AAC	53.9
<b>HERV</b>	MLT1D	TATTAGGAATTGAAAGAGGGAAAGA	Biotin)TCAAAACCACTATAAA AATTACCACAA	TTTAGAGGAAGGATA	55
	ERV1	TTTGTATGGAAGGAGAAATGGTTAG	Biotin)ATACCTCTTCCCCAAAT TTCTTTAT	TTTGTATGGAAGGAGA AATGGTTAG	55
	ERV9	TGTTATTGTTTATTTTTTRGGTTTA	Biotin)TCTTCCTTCTAATAAAT TCATAATCTC	TTATTTTTGAAGTTA	55

We used a PSQ HS 96 Pyrosequencing System (Biotage AB, Uppsala, Sweden), as previously described [64]. Each sample was tested two times for each assay to confirm reproducibility. As a quality control check to estimate the bisulfite conversion efficiency, we placed duplicate genomic DNA samples on each bisulfite conversion plate to estimate the internal plate variation of bisulfite conversion and the Pyrosequencing reaction. We also added universal PCR products amplified from cell-line DNA on each Pyrosequencing plate to check run-to-run and plate-to-plate variation in performing Pyrosequencing reactions. In addition, the pyrogram peak pattern from every sample was checked visually inspected to confirm the quality of the reaction.

#### 4.4 RNA EXPRESSION ANALYSIS

The number of publications focused on detection of L1 RNA is very limited. We choose to analyze the expression of the most representative and well studied subfamilies *AluSx* and *L1HS*. For *L1HS* we chose to analyze ORF1 and ORF2 because they are both involved in the retrotransposition process and to have better results in detecting the full length L1 RNA. The L1 transcriptome consists of a mix of L1 RNA of different sizes including the full length RNA and its variably 5'truncated or spliced. For this reason it is more appropriate to include both sets of ORFs primers. So while the ORF2 set provides an estimation of full length transcripts plus as well as other truncated transcripts, the ORF1 is indicative of the expression level of full-length transcripts.

##### **Real Time PCR**

The best method for detecting L1 RNA expression is Real Time PCR.

Real Time PCR is a technique based on the PCR, which is used to amplify and simultaneously quantify a targeted DNA molecule. The technique is sensitive, relatively simple and easily reproducible [95, 96]. The technique uses fluorescence for the quantification of a gene product, the fluorescence measured reflects the amount of amplified product at each cycle. SYBR green is the simplest detection method for real time PCR. The fact that SYBR green fluoresces 13 times more strongly when bound to dsDNA than when bound to ssDNA [97], allows the levels of dsDNA to be measured throughout the PCR by recording the emission at 521nm (the wavelength at which SYBR green fluoresces).

The fluorescence is detected at each cycle is represented on a curve which has two stages of amplification, an exponential followed by a plateau. The threshold cycle represents the number of cycles which has accumulated enough to be able to develop an amplified fluorescent signal.



The quantification is based on the relationship that exists between the amount of template and the initial value of the threshold cycle obtained during amplification. Real Time PCR allows the quantification of the relative concentrations of gene transcripts by relating the quantity compared to that of a reference housekeeping gene, present in single copy in the genome.

Reverse transcription is the initial step that consist in the production of cDNA from mRNA, and provides a simple way to look at the relative levels of gene transcript.

500 ng total RNA was employed in the reverse transcription (RT) step (25°C for 5 min, 42°C for 30 min, 85°C for 5 min, and then to 4°C) using a random-primed cDNA reaction according to iScript™ Select cDNA Synthesis Kit (Biorad, CA, USA). Real-time polymerase chain reaction (PCR) was carried out in 20 µl of final volume using iQ™ SYBR Green Supermix (Biorad, CA, USA) on an CFX96 Real-Time PCR Detection System (Biorad, CA, USA). Real-time PCR was performed in duplicate, including no-template controls. To prevent variability, we prepared a master mix containing 2.2× volumes of each component in a DNA-free tube, which is subsequently aliquoted in two PCR tubes. Primer sequences and PCR conditions are listed in **Table 6**. Normalization was performed with GAPDH (Glyceraldehyde-3-Phosphate Dehydrogenase) and HPRT1 (Homo Sapiens Hypoxanthine phosphoribosyltransferase 1) as endogenous controls. The relative gene expression was calculated via a  $2^{-\Delta\Delta C_t}$  method [98].

**Table 6.** Primer sequences and PCR conditions for methylation analysis.

Assay	Forward Primer (5' to 3')	Reverse Primer (3' to 5')	Annealing Temperature (°C)
GAPDH	TGCCAAATATGATGACATCAAGAA	GGAGTGGGTGTCGCTGTTG	62.5
HPRT1	ACCCACGAAGTGTTGGATA	AAGCAGATGGCCACAGAACT	62.5
L1Hs ORF1	ATGAGCAAAGCCTCCAAGAA	TGGTTCCATTCTCCACATCA	64.1
L1Hs ORF2	CAGCCGAATTCTACCAGAGG	CCGGCTTTGGTATCAGAATG	64.1
<i>AluSx</i>	GCCTGTAATCCCAGCACTTT	GTAGAGACGGGGTTTCACCA	62.5

## 4.5 STATISTICAL ANALYSIS

### Statistical analysis for the three Studies

We used mixed-effect regression models to evaluate the effects of the exposures on DNA methylation levels of each subfamily, as previously reported. This approach yields a global estimate of the effect on multiple CpGs within each subfamily sequence by modeling correlated data in adjacent CpG sites within each sequence, as well as the measures from duplicate pyrosequencing runs. Mixed-effect models have the advantage over standard methods of using the entirety of the information in the data, thus maximizing statistical power by distinguishing different sources of variance.

We used the following model:

$$[1] \quad Y_{ijk} = \beta_0 + \beta_1 \text{exposure} + \beta_2 X_2 + \cdots + \beta_p X_p + v_{0i} + v_{1ij} \text{position} + \epsilon_{ijk}$$

where  $Y_{ijk}$  represents the methylation level for the  $i$ -th subject, the  $j$ -th position and the  $k$ -th duplicate run ( $i=1, \dots, 40$ ;  $j=1, \dots, m$ , where  $m$  varies depending on the total number of CpG sites measured in the sequence; and  $k=1, 2$ ).  $v_{0i}$  and  $v_{1ij}$  are the random intercept for subject and random slope for CpG position, respectively.  $\beta_0$  is the overall intercept and  $\beta_1$  is the fixed effect which expresses the association between exposure and DNA methylation.  $X_2 \dots X_p$  and  $\beta_2 \dots \beta_p$  represent covariates and their regression coefficients;  $\epsilon_{ijk}$  is the residual term error. Age and smoking were considered *a priori* as possible confounders and therefore included as covariates in all the models of the analysis.

We first fitted a set of models in which DNA methylation was regressed over dichotomous exposure variables (high-exposure vs. low-exposure control groups). In a second

set of models, we evaluated dose-response relationships by regressing DNA methylation over continuous exposure-level variables. To increase goodness-of-fit, all continuous exposure variables were log-transformed. As suggested by Du et al. (Du et al. 2010), we transformed the original methylation measures (bounded between 0 and 100%) in M-values (ranging between  $-\infty$  and  $+\infty$ ) using the following conversion:  $\log_2[\text{meth} (\%5 \text{ mC})/(100 - \text{meth} (\%5 \text{ mC}))]$  in all the models. M-values have been shown to improve homoscedasticity of methylation data and allow for more robust statistical estimates. However, the M-value does not have an intuitive biological meaning and the corresponding model parameters do not have a straightforward interpretation. Therefore, we calculated  $\tau$  as a transformation of the regression coefficients  $\beta$  using the formula  $\tau = (2^\beta - 1) * 100$ , which represents the percent-change of the ratio methylated/unmethylated associated with the exposure. For continuous exposure variables,  $\tau$  was scaled to represent the percent change associated with an increase in exposure from the 25<sup>th</sup> to the 75<sup>th</sup> percentile. We checked regression assumptions by performing diagnostic tests for each model, including the Shapiro-Wilk test for normality of residuals and White test for variance homogeneity of residuals.

We also used mixed-effect regression models to determine whether the correlation between DNA methylation and exposures within each repetitive element family varied as a function of the evolutionary age of the subfamilies. The corresponding model was:

[2]

$$Y_{ijkl} = \beta_0 + \beta_1 \text{exposure} + \beta_2 \text{subfamily\_age} + \beta_3 \text{exposure} * \text{subfamily\_age} + \beta_4 X_4 + \dots + \beta_p X_p + v_{0jk} \text{subfamily} * \text{position} + v_{1l} \text{sample} + \epsilon_{ijkl}$$

where  $Y_{ijkl}$  represents the methylation level for the  $i$ -th subject, the  $j$ -th subfamily, the  $k$ -th position and the  $l$ -th run ( $i=1, \dots, 40$ ;  $j=1, \dots, m$ ; where  $m$  varies depending on the number of

subfamilies evaluated in the family;  $k=1, \dots, n$ , where  $n$  varies depending on the total number of CpG sites measured in the sequence; and  $l=1,2$ ). The random part of the model is composed by the slopes,  $u_{0jk}$  and  $u_{1i}$ , for the interaction between position and subfamily and for subject, respectively. The interaction slope was used to model the existence in the data of positions in common for all the subfamilies. Hence, by using the same labels for the positions in the same common sequences – even if belonging to different subfamilies – and rescaling the others accordingly, the interaction can describe unambiguously to which position and subfamily the measure refers. Finally,  $\beta_0$  is the overall intercept;  $\beta_1$  represents the fixed effect for the exposure,  $\beta_2$  for the age of the subfamily;  $\beta_3$  expresses the interaction between exposure and evolutionary age;  $X_4 \dots X_p$  and  $\beta_4 \dots \beta_p$  are the covariates and their regression coefficients; and  $\epsilon_{ijkl}$  is the residual term error.

A two-sided  $P < 0.05$  was considered statistically significant. All statistical analyses were performed in SAS (version 9.2; SAS Institute Inc., Cary, NC, USA). We used the PROC MIXED procedure to run the mixed-effect models.

### **Statistical analysis for the Beijing Truck Driver Air Pollution Study (BTDAS)**

Due to the particular structure of the data we used a similar model to that proposed by J. Madrigano (2012, American Journal of Epidemiology) with a specific error term structure. DNA methylation at adjacent CpG sites is usually correlated. Therefore, mixed models were used to take full advantage of the information from all measurements in the data and to maximize statistical power by distinguishing between the different sources of variability in the data. Specifically, mixed-effects models with both subject-level and position-specific random effects were used to capture the correlation among measurements within the same subject or the same

location within a promoter region. To test the difference in DNA methylation between truck drivers and office workers we used the following model:

$$Y_{ijk} = \beta_0 + \beta_1(\text{Group}) + \beta_2 X_2 + \dots + \beta_n X_n + u_k + v_i + e_{ijk} \quad (1)$$

where  $Y_{ijk}$  is the measured value of DNA methylation at CpG dinucleotide position  $i$  at day  $j$  of subject  $k$ ;  $\beta_0$  is the overall intercept;  $\beta_1$  is the fixed effect that expresses the difference between the two groups;  $u_k$  is the subject-specific random intercept, which captures the correlation among measurements within the same subject;  $v_i$  is the separate random intercept for each CpG dinucleotide position, which captures the average difference between methylation at that dinucleotide position and the overall mean methylation for the gene; and  $X_2 \dots X_n$  are the covariates for which the model was adjusted. The error term specified in this manner allows the residual variance to vary by position, and for measurements on the different CpG positions taken on the same day to be correlated.

To partition associations between methylation and pollution exposures into a cross-sectional and longitudinal component, the associations of personal PM2.5, personal EC, and ambient PM10 variables (day of the study, 1-day, 2-day, 5-day, or 8-day mean) with DNA methylation variables were evaluated using the following linear mixed effects model:

$$y_{ijk} = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \beta_{n+1}(\overline{Exp_k}) + \beta_{n+2}(Exp_{kj} - \overline{Exp_k}) + u_k + v_i + e_{ijk} \quad (2)$$

$\beta_1$  represents the cross-sectional effects of exposure, describing the manner in which the mean DNA methylation varies with average exposure for a given subject. In contrast,  $\beta_2$  represents the longitudinal effect of exposure since it describes how within-subject changes in the DNA methylation are related to within-subject changes in exposure. The structure of random effects

and the error terms is the same proposed for the model (1). Both models were adjusted for age (continuous), sex (male, female), BMI (continuous), smoking (never, former, current), cigarettes smoked during study time (continuous), work hours/day (continuous), outdoor temperature (continuous) and outdoor humidity (continuous). The average time for outdoor temperature and outdoor humidity were used (one to eight days) to match the average time used for the air particle variables. As a primary analysis we fitted the second model using all participants; a second analysis was then conducted, evaluating the associations with air pollution in office workers and truck drivers separately. We considered all tests with two-sided and significant alpha level of 0.05. All effect estimates ( $\beta$ ) and their 95% confidence intervals (CI) are presented as percent changes per interquartile range (IQR) change of each exposure.

It is well known the strict relationship between DNA methylation and gene expression, therefore, the mixed models taking into account the gene expression data were used to understand whether the results about the DNA methylation are consistent.

The gene expression data ( $C_T$ ) were transformed according to the pipeline of Bio-Rad, then the outcomes, *AluSX*, ORF1 and ORF2, were singularly and graphically analyzed and checked the quality of the data: after a logarithm transformation of the data, to get their normality, the outliers were discarded.

The models used for the gene expression data were mixed-effects model with subject-level random effect to capture the correlation among measurements within the same subject.

The following model was used to test the difference in gene expression between the two groups, truck drivers and office workers:

$$Y_{jk} = \beta_0 + \beta_1(\text{Group}) + \beta_2 X_2 + \dots + \beta_{n-1} X_{n-1} + u_k + e_{jk} \quad (3)$$

where  $Y_{jk}$  is the logarithm of the measured value of the gene expression at day  $j$  of subject  $k$ ;  $\beta_0$  is the overall intercept;  $\beta_1$  is the fixed effect that express the difference between the two groups,  $u_k$  is the subject-specific random intercept and  $X_2, \dots, X_{n-1}$  are the covariates for which the model was adjusted. We assume that  $u_k$  and  $e_{jk}$  are independent and normally distributed with variance  $\sigma_k^2$  and  $\sigma_e^2$ , respectively.

The associations between gene expression and the same pollution exposures of model (2) can be split into a cross-sectional and a longitudinal component using the following linear mixed-effects model:

$$Y_{jk} = \beta_0 + \beta_2 X_2 + \dots + \beta_{n-1} X_{n-1} + \beta_{n+1} (\overline{Exp_k}) + \beta_{n+2} (Exp_{jk} - \overline{Exp_k}) + u_k + e_{jk} \quad (4)$$

where  $\beta_{n+1}$  is the cross sectional effect of the exposure, which exponentiated describes how the mean of the gene expression varies with average exposure for a given subject and  $\beta_{n+2}$  is the longitudinal effect that exponentiated expresses how within-subject changes in the gene expression are related to within-subject changes in exposure. The structure of random effects and the error terms is the same proposed for the model (3).

Model (3) and (4) were adjusted for age (continuous), sex (male/female), BMI (continuous) smoking status (never, former, current), cigarettes smoking during study time (continuous), working hours/day (continuous) and outdoor temperature (continuous). The average time for outdoor temperature was used (one to eight days) to match the average time used for the air particle variables.

As in the DNA methylation model a primary analysis of the forth model was fitted using all participants and in a second step was evaluated the associations with air pollution in the two groups separately. All effect estimates ( $\beta$ ), in model (3) and (4), were exponentiated to get



predictions on the original scale of the outcome variable and all coefficients and their 95% confidence interval (CI) were presented as percent changes per unit change of each exposure. We considered all tests with two-sided and significant alpha level of 0.05.

Pearson's and Spearman's correlation coefficients between the average of the gene methylation and the average of the gene expression were computed in order to understand whether, respectively, a linear or a monotonous relationship existed between them.

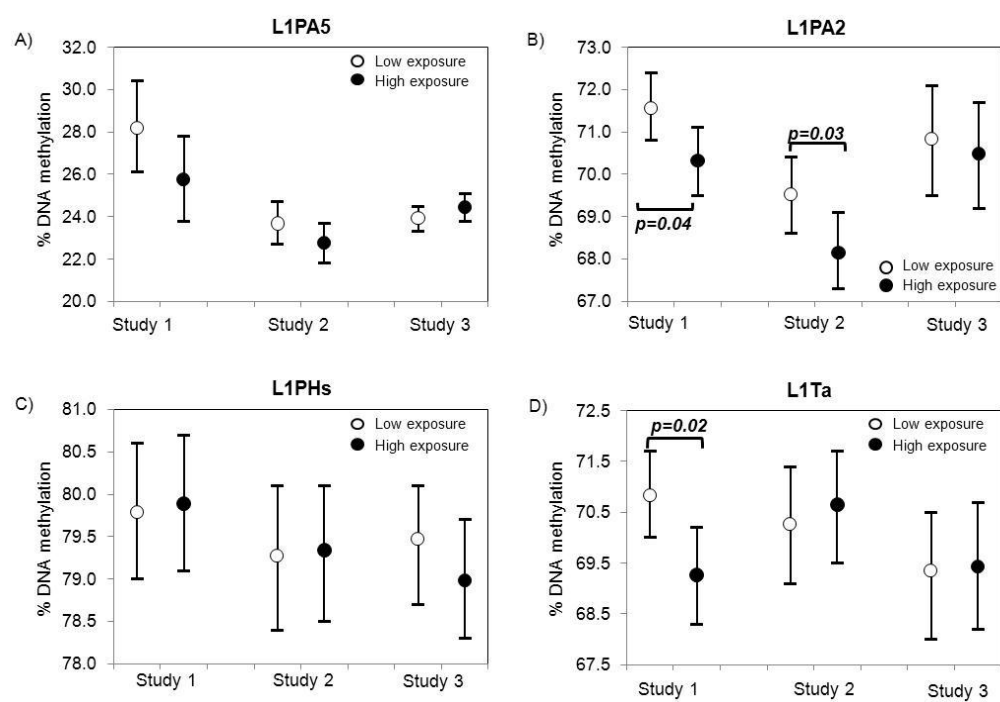
## 5. RESULTS

### 5.1 METHYLATION ANALYSIS FROM THE THREE STUDIES

#### Levels of DNA methylation of repetitive elements subfamilies by exposure group

We first examined the effect of environmental exposures on DNA methylation in repetitive element subfamilies, by contrasting high- vs. low-exposure groups in analysis adjusted for age and smoking. Among LINE-1 subfamilies, L1PA2 showed significantly lower DNA methylation both in steel workers highly-exposed to metal-rich PM<sub>10</sub> in Study 1 (mean differences=-1.2%,  $P=0.04$ ) and in gas-station attendants in Study 2 (mean differences=-1.3%,  $P=0.03$ ) (**Figure 1B**). L1Ta, the youngest of the LINE-1 showed in Study 1 significantly decreased DNA methylation in steel workers with high exposure to metal-rich PM compared to the low-exposed group (mean differences=-1.5%,  $P=0.02$ ) (**Figure 1D**). Neither L1PA5—the oldest LINE-1 subfamily in this study—nor L1Hs, a relatively young subfamily, showed significant DNA methylation differences between low and high exposure groups in any of the three studies (**Figure 1A** and **1C**).

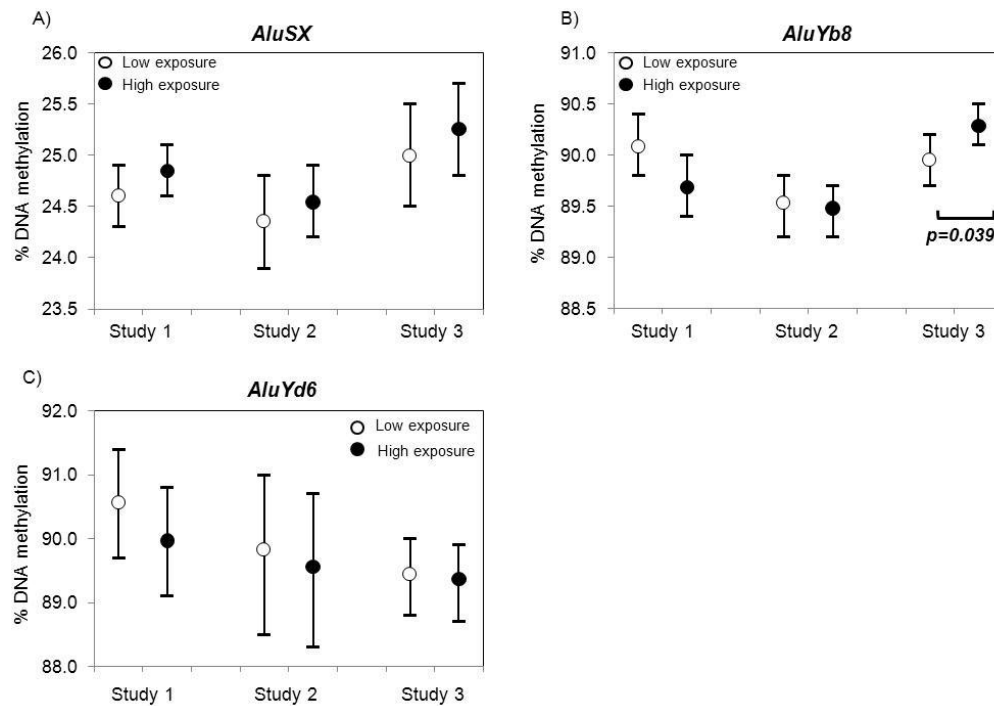
**Figure 1.** DNA methylation differences in LINE-1 subfamilies between low and high exposure groups.



*AluYb8* –a relatively younger *Alu* –showed in Study 3 significantly increased DNA methylation in truck drivers with high EC exposure compared to indoor office workers (mean difference=0.4%,  $P=0.039$ ) (**Figure 2B**). Neither *AluSx* – the oldest *Alu* subfamily in this study – nor *AluYd6* – the youngest *Alu* subfamily in this study –showed significant differences in DNA methylation in any of the three studies (**Figure 2A** and **2C**).

None of the HERV subfamilies showed significant DNA methylation differences between high and low-exposure groups in any of the three studies.

**Figure 2** DNA methylation differences of *Alu* subfamilies between low and high exposure group.



### Correlations of DNA methylation in repetitive elements subfamilies with exposure levels

Among LINE-1 subfamilies, all the significant correlations of exposure levels with DNA methylation across studies were observed in the two oldest subfamilies, i.e., L1PA5 and L1PA2. In Study 1, DNA methylation of L1PA5 and L1PA2 were negatively correlated with the levels of exposure to metal-rich PM<sub>10</sub> ( $\tau=-15.2$ ,  $P=0.02$  and  $\tau=-5.8$ ,  $P=0.03$ , respectively; **Table 1**). In Study 2, L1PA2 methylation showed a negative correlation with air benzene exposure ( $\tau=-4.3$ ,  $P=0.01$ ). In Study 3, L1PA5 methylation showed a negative correlation with EC exposure ( $\tau=5.6$ ,  $P=0.01$ ). In all the three studies, there were no or marginal correlations between methylation of *Alus* and HERVs with the levels of exposure.

**Table 1** Dose-response relationship between levels of personal exposure to air pollutants and DNA methylation in the three study populations.

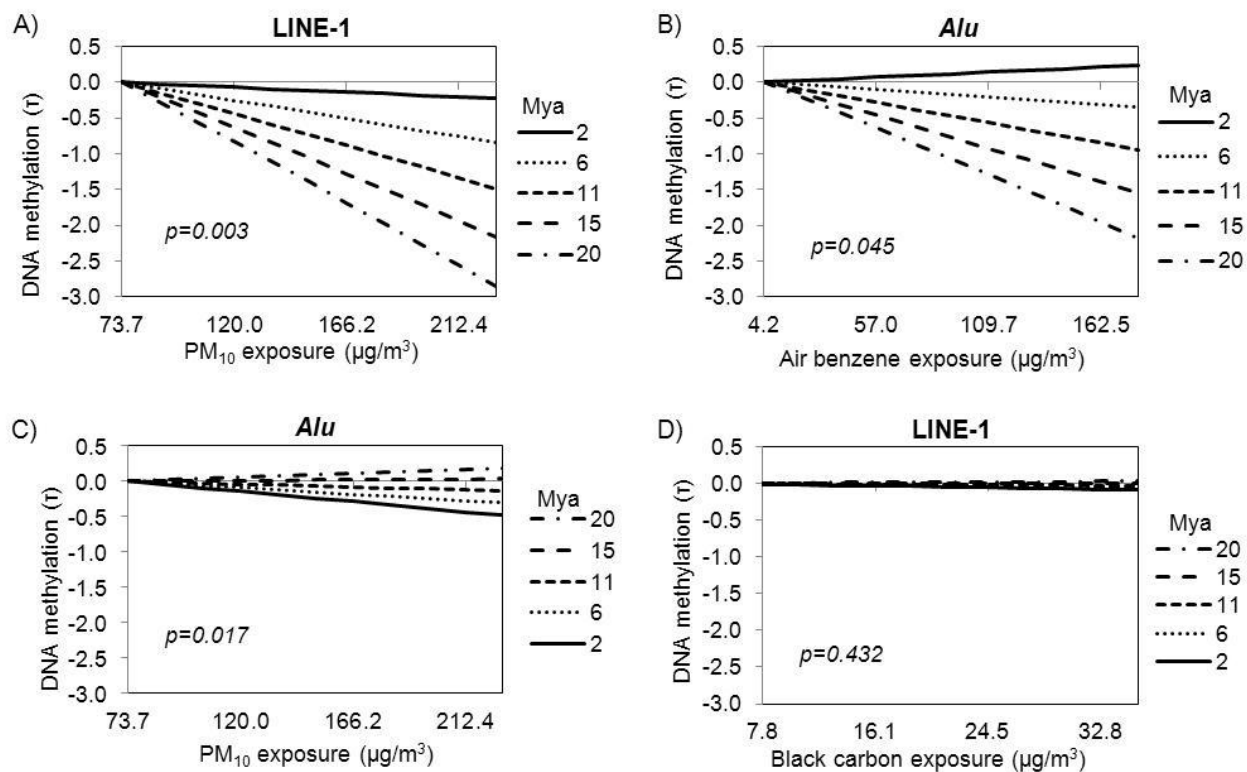
Repetitive Element		Study 1 Associations with metal-rich particulate matter (PM <sub>10</sub> )		Study 2 Associations with air benzene		Study 3 Associations with traffic- derived elemental carbon	
Family	Subfamily	$\tau$	p value	$\tau$	p value	$\tau$	p value
LINE-1	L1PA5	-15.2	<b>0.02</b>	-2.3	0.28	5.6	<b>0.01</b>
	L1PA2	-5.8	<b>0.03</b>	-4.3	<b>0.01**</b>	-2.3	0.51
	L1Hs	1.4	0.68	1.0	0.60	-2.7	0.19
	L1Ta	-4.2	0.17	0.7	0.74	1.6	0.68
<i>Alu</i>	<i>AluSx</i>	0.0	1.00	1.0	0.21	0.0	1.00
	<i>AluYb8</i>	-3.8	0.11	-0.4	0.74	2.4	0.14
	<i>AluYd6</i>	-6.6	0.31	-4.1	0.37	4.0	0.29
HERV	MLT1D	-47.6	0.16	54.6	0.13	-41.8	0.09
	ERV1	0.4	0.78	0.6	0.56	0.0	0.99
	ERV9	1.9	0.45	1.4	0.42	0.5	0.83

### Evolutionary age of subfamilies and DNA methylation sensitivity to the exposures

To determine whether repetitive element sensitivity to environmental exposures was dependent on evolutionary age, we used an interaction analysis that modeled the correlation between environmental exposure levels and DNA methylation as a function of evolutionary age. **Figure 3** presents a graphical representation of the results for those repetitive element families and type of exposures that showed significant correlations with exposure levels, as described in the section above. To illustrate the model results, we present graphical representations of the slopes for the effects of the exposures on DNA methylation, estimated at arbitrary evolutionary ages (expressed in Mya) selected to be equally spaced across the age ranges of the subfamilies analyzed.

In Study 1, the negative effect of PM<sub>10</sub> exposure on LINE-1 methylation was significantly pronounced as evolutionary age increased ( $\tau$  for interaction = -0.6;  $P=0.003$ ) (**Figure 3A**). In Study 2, the negative effect of airborne benzene in LINE-1 DNA methylation was also progressively stronger with increasing evolutionary age ( $\tau$  for interaction = -0.2;  $P=0.045$ ) (**Figure 3B**). In Study 1, we observed a positive significant interaction between PM<sub>10</sub> exposure and evolutionary age in predicting *Alu* family methylation ( $\tau=0.2$  and  $P=0.017$ ) (**Figure 3C**). However, as shown in the slopes in **Figure 3**, the differences of the effects of PM<sub>10</sub> across different evolutionary ages were much less pronounced for *Alu* methylation (**Figure 3C**) compared to LINE-1 methylation (**Figure 3B**). The effects of elemental carbon exposure on LINE-1 methylation did not show significant variations by evolutionary age (**Figure 3D**). No interactions were found between evolutionary ages and exposure levels in determining HERV methylation.

**Figure 3.** Interaction of evolutionary ages (Millions year ago, Mya) with exposure in relation to difference in repetitive element DNA methylation.



## 5.2 METHYLATION ANALYSIS FROM THE BEIJING TRUCK DRIVER AIR POLLUTION STUDY (BTDAS)

### Level of DNA methylation of repetitive elements subfamilies by exposure group

We first examined the effect of environmental exposures on DNA methylation in repetitive element subfamilies, by contrasting truck drivers vs. office workers groups. Comparing truck drivers with indoor workers allows examination of the long-term PM<sub>2.5</sub> effects because of the substantial difference in their usual PM<sub>2.5</sub> exposure levels.

The two groups were matched by sex and smoking habits (never, former, current smoker). The analysis were adjusted for age, number of smoked cigarettes and work hours per day. Among *Alu* subfamilies, *AluSX* -the oldest between the *Alu* subfamilies -showed significantly lower DNA methylation in the truck drivers group ( $\beta=-0.19$ ,  $P=0.02$ ). Among HERV subfamilies, *MLT1d* - the oldest between the HERV subfamilies- showed significantly lower DNA methylation in the truck drivers group ( $\beta=-0.64$ ,  $P=0.01$ ). We did not observe any significant difference between groups in the *Alu* and HERV youngest subfamilies. None of the LINE-1 subfamily in this study showed significant DNA methylation differences between truck drivers and office workers groups (**Table 2**).



**Table 2:** DNA methylation differences in repetitive element subfamilies between groups.

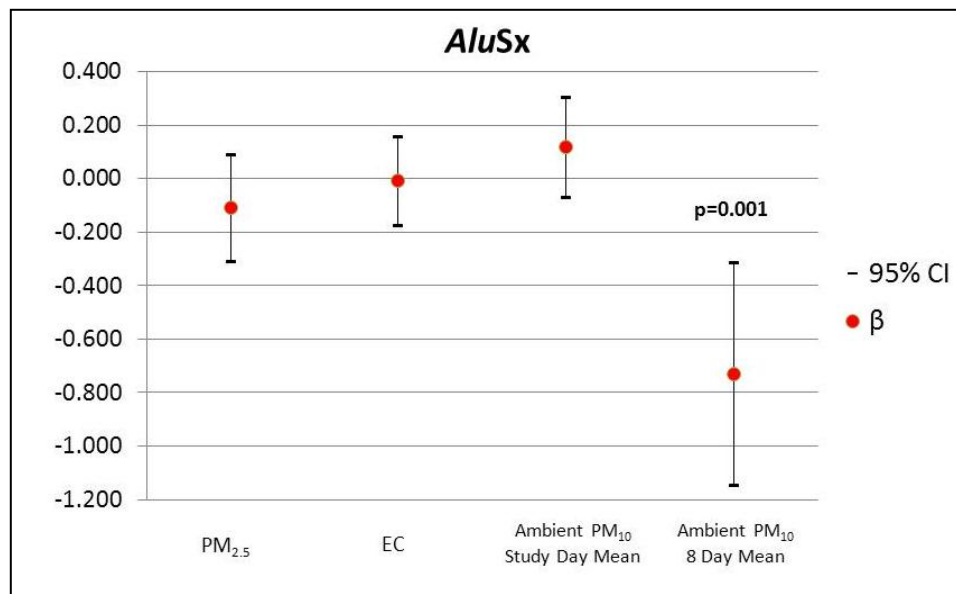
Repetitive element subfamily	Unadjusted							Adjusted *						
	Office Workers		Truck Drivers		Difference between Groups			Office Workers		Truck Drivers		Difference between Groups		
	<i>Mean</i>	<i>SE</i>	<i>Mean</i>	<i>SE</i>	$\beta$	<i>SE</i>	<i>p-value</i>	<i>Mean</i>	<i>SE</i>	<i>Mean</i>	<i>SE</i>	$\beta$	<i>SE</i>	<i>p-value</i>
AluSX	28.61	-3.9	28.62	-3.9	0.02	-0.08	0.83	28.45	-3.87	28.26	-3.87	-0.19	-0.08	<b>0.02</b>
AluYb8	90	-1.87	89.85	-1.87	-0.15	-0.11	0.17	90.93	-1.29	90.83	-1.32	-0.09	-0.14	0.5
Alud6	91.62	-0.68	91.6	-0.68	-0.02	-0.18	0.93	92.83	-1.04	92.92	-1.11	0.08	-0.22	0.7
L1PA5	25.05	-0.24	24.44	-0.24	-0.61	-0.34	0.08	26.45	-1.68	26.13	-1.86	-0.32	-0.43	0.45
L1PA2	70.19	-7.42	70.02	-7.42	-0.17	-0.17	0.31	70.18	-7.46	70.12	-7.47	-0.05	-0.21	0.8
L1HS	82.18	-1.84	81.97	-1.84	-0.21	-0.2	0.29	83.25	-2.05	83.14	-2.1	-0.12	-0.25	0.64
L1Ta	74.7	-7.62	74.78	-7.62	0.08	-0.16	0.63	74.28	-7.66	74.28	-7.67	0	-0.2	1
MLT1d	95.87	-0.49	95.69	-0.49	-0.17	-0.21	0.41	93.87	-1.04	93.22	-1.15	-0.64	-0.26	<b>0.01</b>
HERVI	25.5	-0.1	25.43	-0.1	-0.07	-0.14	0.61	26.78	-0.72	26.75	-0.78	-0.03	-0.16	0.86
HERV9	54.95	-0.16	54.8	-0.16	-0.15	-0.22	0.5	56.39	-1.3	56.55	-1.4	0.16	-0.28	0.56

### Cross-sectional effect of air particles on DNA Methylation

To evaluate the dose-response relationship between specific exposures and DNA methylation, we run a cross-sectional analyses. We adjusted for age, sex, BMI, smoking status, number of cigarettes smoked on the study day, work hours/day, appropriate outdoor temperature and humidity (i.e. temprature and humidity averaged over the same time window as the air particle exposure). We confirmed that the most affected subfamilies were *AluSx* and *MLT1d*.

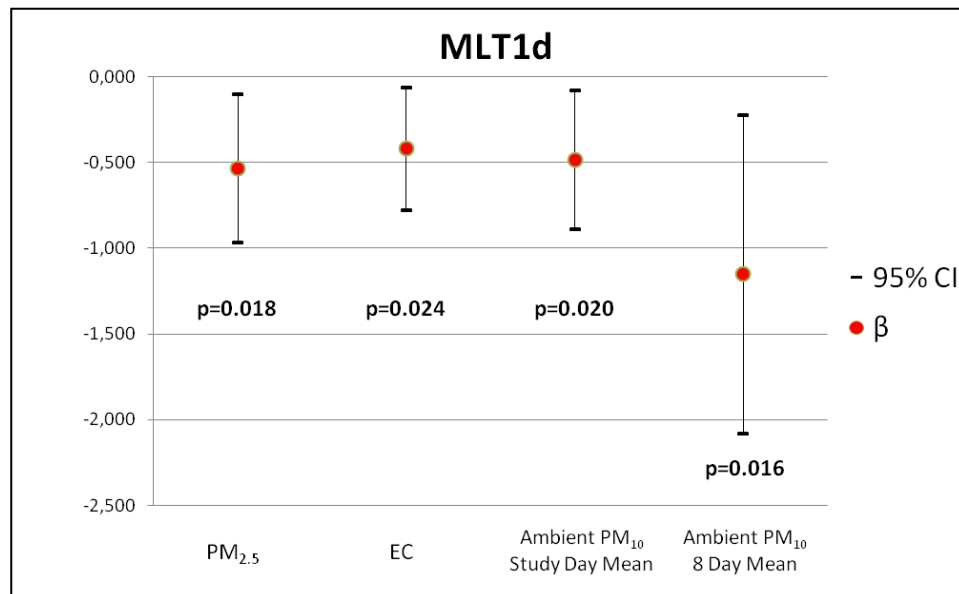
Among *Alu* subfamilies, again the oldest *AluSx*, showed significantly lower DNA methylation in relation to  $PM_{10}$  8-day mean (i.e., average of the 8 days before the examination) exposure in all the subjects ( $\beta=-0.226$ ,  $P=0.047$ ) (**Figure 3**).

**Figure 3.** Cross -sectional effect of air particles on *AluSx* DNA methylation.



Among HERV subfamilies, a significant lower methylation of the oldest MLT1d was observed for all the subjects in relation to exposure to PM<sub>2.5</sub> ( $\beta=-0.665$ ,  $P=0.002$ ), EC ( $\beta=-0.498$ ,  $P=0.008$ ), ambient PM<sub>10</sub> study day mean ( $\beta=-0.582$ ,  $P=0.005$ ) and ambient PM<sub>10</sub> 8-day mean ( $\beta=-0.961$ ,  $P=0.018$ ) (**Figure 4**).

**Figure 4.** Cross -sectional effect of air particles on MLT1d DNA Methylation.



### 5.3 EXPRESSION ANALYSIS FROM THE BEIJING TRUCK DRIVER AIR POLLUTION STUDY (BTDAS)

#### Expression level of repetitive elements subfamilies by exposure group

In order to study the dose-response relationship we examined the effect of environmental exposures on the expression levels of *AluSx* and *L1HS* (measured as ORF1 and ORF2 expression) subfamilies, by contrasting truck drivers vs. office workers groups. The two groups were matched by sex and smoking habits (never, former, current smoker). The analysis were adjusted for age, gender, BMI, smoking status, number of cigarettes smoked on the study day, work hours/day, appropriate outdoor temperature (i.e. temperature averaged over the same time window as the air particle exposure variables) and date.

We observed a slightly but not statistically significant increased expression level in the truck drivers group for all the analyzed regions.

**Table 3:** Expression differences in repetitive element subfamilies between groups.

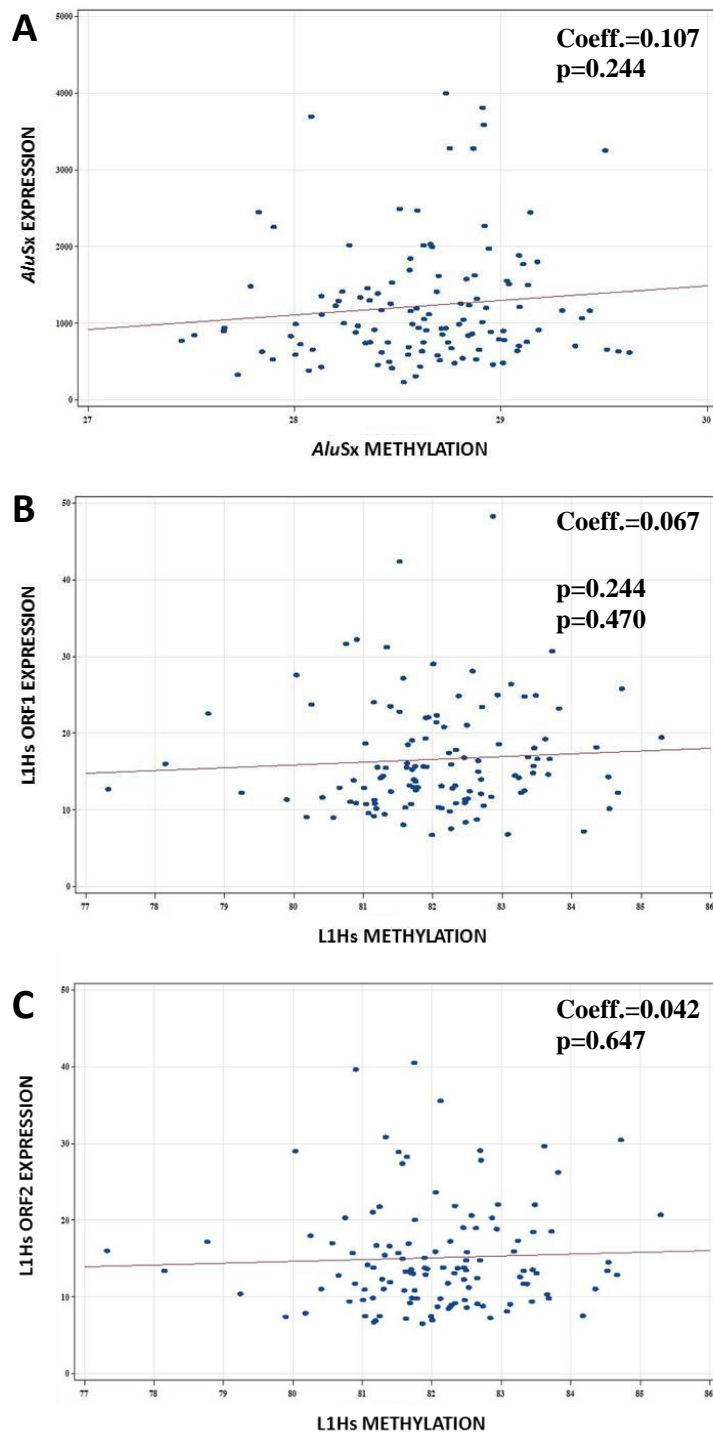
Expression	All subjects		Office Workers		Truck Drivers		Difference between Groups		
	<i>Mean</i>	<i>SE</i>	<i>Mean</i>	<i>SE</i>	<i>Mean</i>	<i>SE</i>	<i>Beta</i>	<i>SE</i>	<i>p-value</i>
<i>Alu Sx</i>	1247.251	70.366	1173.484	82.701	1319.736	113.364	1.606	-8.502	0.892
ORF1	16.594	0.638	16.699	0.93	16.488	0.877	0.04	-9.305	0.996
ORF2	15.216	0.627	15.185	0.911	15.249	0.862	4.466	-12.451	0.624

### **Correlation between repetitive element methylation and expression levels**

Pearson's correlation analysis was performed to evaluate the correlations between: i) the methylation and the expression levels of *AluSx*, ii) the methylation level of L1HS and the expression level of ORF1 and iii) the methylation level of L1HS and the expression level of ORF2.

We observed a slightly but not statistically significant positive correlation between the methylation and the expression levels in all the three analysis (**Figure 5**). Both the coefficients were not significantly different from zero, therefore there were no empirical evidence of linear or monotonous relationship between the methylation and gene expression.

**Figure 5.** Correlation between repetitive element methylation and expression levels.



## 6. DISCUSSION

DNA methylation of repetitive elements has been extensively studied in relation to environmental exposures and human disease. Nonetheless, most if not all of the previous studies have investigated one single sequence in one or at most two repetitive element subfamily. The present work is based on a comprehensive methylation analysis of 10 repetitive elements subfamilies with different ages (old, intermediate, and young) depending on the time when they appeared in the human genome. We examined three groups of participants with well-characterized exposure including steel workers with exposure to PM<sub>10</sub>; gas-station attendants exposed to air benzene; and truck drivers exposed to EC (BTDAS). We evaluate sensitivity of DNA methylation of the same 10 subfamilies in a more wide population of 120 individuals (Beijing Truck Driver Air Pollution Study, BTDAS) with a well characterized personal exposure levels of PM<sub>2.5</sub> and ambient PM<sub>10</sub>. We found that effects on DNA methylation of individual repetitive elements subfamilies were specific to the exposure type. We showed that some of the effects identified were dependent on the subfamily evolutionary age.

We further evaluate the expression of the most representative and well studied subfamilies *AluSx* and *L1HS* in 120 individuals from BTDAS cohort and we found no difference in the expression levels between the two groups. There was no correlation between the methylation and expression levels of *AluSx* and *L1HS* repetitive elements.

Methylation of individual sequences in the repetitive element families LINE-1, *Alu*, and HERV has been already investigated in relation to environmental exposures such as PM<sub>10</sub>, black carbon, and persistent organic pollutants [73, 99]. However, due to the similarity of the sequences and

the difficulty in designing primers for specific subfamilies, most of previous studies have analyzed only one single subfamily (i.e., L1Hs for the LINE-1 family and *AluSx* for the *Alu* family) and no study has yet focused on the evaluation of multiple subfamilies. In the present study, we found significant differences in methylation level of specific repetitive element subfamilies in both the analysis using exposure group (high vs. low) and the dose-response analysis using continuous exposure levels to metal-rich PM, air benzene, or EC. However in the three studies, not all the effects on DNA methylation were consistently found in both the group and dose-response analysis. For instance, L1Ta showed a significant difference in high-exposed steel-workers in Study 1 the group analyses that was not confirmed in the dose-response analysis using continuous PM<sub>10</sub> levels. Similarly, in the *Alu* family, DNA methylation of the intermediate-age *AluYb8* sequence showed a significant difference in the high-exposure group of truck drivers in Study 3, which was not confirmed in the dose-response analysis using continuous EC levels. The discrepancies of DNA methylation levels found in both the group and dose-response analysis in the three studies, might be explained by the small number of individuals in each group. It is worth noting that in the Beijing study with an augmented sample size, we didn't find those inconsistencies any more. It is also worth noting that even in the cases with discordance of statistical significance, group and dose-response analysis were concordant in showing similar directions for the exposure-related methylation differences.

Repetitive element subfamilies were inserted in the host human genome at different evolutionary ages. To provide more stable estimates of the general effects of air pollutants, as well as to elucidate the biological bases of the differences of effects within repetitive element family, we investigated the subfamily evolutionary age in the correlations between subfamily methylation



and exposures. We observed that the effects of air pollutants on repetitive-element methylation – particular in LINE-1 subfamilies – were significantly affected by the age of subfamilies. The interaction analysis of environmental exposure and ages of repetitive elements subfamilies suggested that DNA methylation from older LINE-1 subfamilies may be more vulnerable to environmental exposure than in younger subfamilies.

The different susceptibility we observed in relation to environmental pollutants may be first explained by considering the sequence variation and the GC contents between the subfamilies. Subfamilies with older evolutionary age have lower CpG content due to higher substitution rates. As a measure of CpG density, we calculated the ratio between observed and expected CpG content (CpG[o/e]) for each subfamily. Our data confirmed that the subfamily age was inversely correlated with DNA methylation levels in the CpG sites. Also, DNA methylation of those CpG sites was positively correlated with the ratio of CpG[o/e]. These findings show that older subfamilies have lower CpG density and are prone to have lower DNA methylation. This also supports the concept that each repetitive element family has different patterns of DNA methylation, which might reflect varying degrees of regulation and help explaining the different responses to environmental exposures.

Another possible explanation for the exposure-related differences in DNA methylation observed in the present study could be found in the genomic position of repetitive elements in the human genome. Repetitive element families show different insertional preference in the human genome; for instance, LINE-1s are frequently inserted in AT rich regions as TTTT/A is the site to prime reverse transcription [99]. *Alus* and ERVs are more likely inserted into GC rich regions, i.e. in regions near genes or gene-related features such as CpG islands [100]. Due to the functional

relationship with repetitive elements and their different structural characteristics, surrounding regions might be differentially sensitive to the environmental exposures.

Our findings are consistent with activation of pollutant-specific biological pathways, which may in turn result in signature differences in DNA methylation in specific repetitive element subfamily. DNA methyltransferases (DNMTs) play a fundamental role in the methylation process by transferring the methyl group from S-adenosyl-methionine (SAM) to the C5 position of the pyrimidine ring of cytosine. DNMTs are environmentally sensitive [48] and may represent vulnerable targets in the biological process linking pollutant exposures to DNA methylation. Specific DNMT isoforms show different sensitivity to the environment, as potentially each pollutant might target one or a combination the several DNMT isoforms. DNMT isoforms also have different activity in the methylation of individual repetitive element subfamilies [101]. For instance an *in vitro* study, LINE-1 sequences are preferentially methylated by the DNMT3B1, DNMT3B2, and DNMTΔ3B isoforms, which however do not produce any methylation on *AluYb8*. Taken together, these data indicate that different susceptibility of DNMTs to environmental exposures could modify their subfamily-specific activity on DNA methylation.

This is the first study that analyzed the methylation a wide number of repetitive element subfamilies in association with three different exposure. Significant differences in methylation level were found also in repetitive element subfamilies that have never been analyzed in previous epidemiological studies. Therefore previous studies that evaluated the global methylation just considering the specific subfamilies of L1 HS and *AluSx* might have missed important associations with environmental exposures. Studying an augmented panel of repetitive element subfamilies might help to identify novel associations in relation to environmental exposures.

However valuating the methylation of a wide number of repetitive element subfamily might be labor intense and cost expensive. Therefore this study might be taken into account for future epidemiological studies that want to focus on the evaluation of global methylation in association to specific environmental exposure.

This is also the first epidemiological study that analyzed the correlation between methylation and expression of repetitive elements in response to environmental exposure. Although we did not find any significant difference in expression levels. The reason of this lack of correlation could be due to methylation-independent mechanisms. Retrotransposition could be impaired by methylation-independent mechanisms such as Small interference RNAs (siRNAs), premature transcript termination and transcriptional elongation defects.

Small interference RNAs (siRNAs) produced by the LINE-1 5' antisense promoter have been shown to inhibit its retrotransposition by 50% [102]. RNA silencing refers to a particular collection of phenomena in which small regulatory noncoding RNAs of 19–28 nucleotides (nt), derived from double-stranded RNAs (dsRNAs), induce gene silencing through various mechanisms, including blocking productive translation of messenger RNAs by RNAi, or transcriptional gene silencing by induction of heterochromatization of specific target DNAs.

The 5' UTR of human L1 is approximately 900 nt and contains an internal promoter that drives transcription at or near position +1. Notably, a ubiquitously active antisense promoter (ASP) has been identified in the 5' UTR, giving rise to many chimeric transcripts of adjacent cellular genes. Thus, it is possible that dsRNA with the L1 sequence is produced by bidirectional transcription of the 5' UTR. dsRNA derived from the human L1 5' UTR can be processed into siRNAs and reduce retrotransposition in cell culture [102].

Another mechanism that could explain poor L1 expression is the premature transcript termination. The A-rich coding strand of the human LINE-1 contains multiple functional canonical and non canonical polyadenylation (poly(A)) signals, resulting in truncation of full-length transcripts by premature polyadenylation. The widespread presence of these poly(A) signals suggests that they have a conserved function, perhaps limiting, or regulating, LINE-1 retrotransposition.

The last mechanism that regulates L1 and *Alu* expression is primarily due to inadequate transcriptional elongation. Because L1 is an abundant and broadly distributed mobile element, the inhibition of transcriptional elongation by L1 might profoundly affect expression of endogenous human genes. It has been shown that poor expression of L1 elements results from the inability of RNA polymerase to elongate efficiently through L1 coding sequence

The present study has a number of limitations. The small sample size of only 40 participants from each of the three studies might have limited the power to detect exposure-related differences. We think that we might have overcome this limit with the analysis of the BTDA Study. Despite designing PCR primers on highly homologous sequence regions between subfamilies, our assays might have missed some copies of each subfamily due to the sequence variations inherent repetitive element subfamilies. In addition, because of the general characteristics of sodium-bisulfite conversion, we could not distinguish between CpG to TpG mutation and cytosine methylation in CpG sites. Separate genomic sequencing would be necessary to identify *bona fide* cytosine methylation in CpG sites. However, mutations in somatic non-malignant cells are rare events and are thus not expected to cause meaningful deviations in the measured methylation levels. The selection of repetitive elements in this study

was limited to representative sequences with different evolutionary ages. Future studies are needed including a larger number of repetitive element subfamilies. Nonetheless – to the best of our knowledge – the present study includes the most comprehensive selection to date of subfamilies ever examined in relation with environmental exposure.

This study has also a number of strengths that support the validity of the results. We conducted a comprehensive DNA methylation analysis of repetitive element subfamilies of different evolutionary ages within three distinct families. We used a highly *quantitative* bisulfite-PCR-*pyrosequencing* approach for DNA methylation analysis, which is the gold standard for DNA methylation analyses in short (up to 80–100 bp) sequences. Finally, we evaluated three different airborne pollutants, whose exposure was measured through directly-measured or estimated levels at the personal level.

In conclusion, the present study on DNA methylation of ten repetitive element subfamilies showed family- and subfamily-specific effects from three distinct air pollutants. We also showed that sensitivity to environmental exposures is dependent on the evolutionary ages of the repetitive elements. Our results provide better understanding of the effects of the exposures on methylation of repetitive element subfamilies and may help to elucidate the role of repetitive in response to environmental risk factors related to human health and disease.

## 7. REFERENCES

1. Beauregard, A., M.J. Curcio, and M. Belfort, *The take and give between retrotransposable elements and their hosts*. Annu Rev Genet, 2008. **42**: p. 587-617.
2. Goodier, J.L. and H.H. Kazazian, Jr., *Retrotransposons revisited: the restraint and rehabilitation of parasites*. Cell, 2008. **135**(1): p. 23-35.
3. Han, J.S. and J.D. Boeke, *A highly active synthetic mammalian retrotransposon*. Nature, 2004. **429**(6989): p. 314-8.
4. Kazazian, H.H., Jr., *Mobile elements: drivers of genome evolution*. Science, 2004. **303**(5664): p. 1626-32.
5. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
6. Sverdlov, E.D., *Retroviruses and primate evolution*. Bioessays, 2000. **22**(2): p. 161-71.
7. Brouha, B., et al., *Hot L1s account for the bulk of retrotransposition in the human population*. Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5280-5.
8. Goodier, J.L., et al., *LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex*. Mol Cell Biol, 2007. **27**(18): p. 6469-83.
9. Moran, J.V., et al., *High frequency retrotransposition in cultured mammalian cells*. Cell, 1996. **87**(5): p. 917-27.
10. Luan, D.D., et al., *Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition*. Cell, 1993. **72**(4): p. 595-605.
11. Feng, Q., et al., *Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition*. Cell, 1996. **87**(5): p. 905-16.
12. Dewannieux, M., C. Esnault, and T. Heidmann, *LINE-mediated retrotransposition of marked Alu sequences*. Nat Genet, 2003. **35**(1): p. 41-8.
13. Kramerov, D.A. and N.S. Vassetzky, *Short retroposons in eukaryotic genomes*. Int Rev Cytol, 2005. **247**: p. 165-221.
14. Eickbush, T.H., *Transposing without ends: the non-LTR retrotransposable elements*. New Biol, 1992. **4**(5): p. 430-40.
15. Cordaux, R., et al., *Estimating the retrotransposition rate of human Alu elements*. Gene, 2006. **373**: p. 134-7.
16. Jurka, J. and T. Smith, *A fundamental division in the Alu family of repeated sequences*. Proc Natl Acad Sci U S A, 1988. **85**(13): p. 4775-8.
17. Giordano, J., et al., *Evolutionary history of mammalian transposons determined by genome-wide defragmentation*. PLoS Comput Biol, 2007. **3**(7): p. e137.
18. Khan, H., A. Smit, and S. Boissinot, *Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates*. Genome Res, 2006. **16**(1): p. 78-87.
19. Cordaux, R. and M.A. Batzer, *The impact of retrotransposons on human genome evolution*. Nat Rev Genet, 2009. **10**(10): p. 691-703.

20. Hough-Evans, B.R., R.J. Britten, and E.H. Davidson, *Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo*. Dev Biol, 1988. **129**(1): p. 198-208.
21. Batzer, M.A., et al., *Structure and variability of recently inserted Alu family members*. Nucleic Acids Res, 1990. **18**(23): p. 6793-8.
22. Ehrlich, M., et al., *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*. Nucleic Acids Res, 1982. **10**(8): p. 2709-21.
23. Matera, A.G., U. Hellmann, and C.W. Schmid, *A transpositionally and transcriptionally competent Alu subfamily*. Mol Cell Biol, 1990. **10**(10): p. 5424-32.
24. Jordan, I.K., *Evolutionary tinkering with transposable elements*. Proc Natl Acad Sci U S A, 2006. **103**(21): p. 7941-2.
25. Walsh, C.P., J.R. Chaillet, and T.H. Bestor, *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation*. Nat Genet, 1998. **20**(2): p. 116-7.
26. Slotkin, R.K. and R. Martienssen, *Transposable elements and the epigenetic regulation of the genome*. Nat Rev Genet, 2007. **8**(4): p. 272-85.
27. Medstrand, P., et al., *Impact of transposable elements on the evolution of mammalian gene regulation*. Cytogenet Genome Res, 2005. **110**(1-4): p. 342-52.
28. Guil, S. and M. Esteller, *DNA methylomes, histone codes and miRNAs: tying it all together*. Int J Biochem Cell Biol, 2009. **41**(1): p. 87-95.
29. El-Osta, A., *DNMT cooperativity--the developing links between methylation, chromatin structure and cancer*. Bioessays, 2003. **25**(11): p. 1071-84.
30. Jaenisch, R., *DNA methylation and imprinting: why bother?* Trends Genet, 1997. **13**(8): p. 323-9.
31. Jones, P.A., et al., *De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines*. Proc Natl Acad Sci U S A, 1990. **87**(16): p. 6117-21.
32. Orphanides, G. and D. Reinberg, *A unified theory of gene expression*. Cell, 2002. **108**(4): p. 439-51.
33. Miranda, T.B. and P.A. Jones, *DNA methylation: the nuts and bolts of repression*. J Cell Physiol, 2007. **213**(2): p. 384-90.
34. Latham, J.A. and S.Y. Dent, *Cross-regulation of histone modifications*. Nat Struct Mol Biol, 2007. **14**(11): p. 1017-24.
35. Martens, J.H., et al., *The profile of repeat-associated histone lysine methylation states in the mouse epigenome*. EMBO J, 2005. **24**(4): p. 800-12.
36. Kouzarides, T., *SnapShot: Histone-modifying enzymes*. Cell, 2007. **128**(4): p. 802.
37. Okitsu, C.Y. and C.L. Hsieh, *DNA methylation dictates histone H3K4 methylation*. Mol Cell Biol, 2007. **27**(7): p. 2746-57.
38. Schubeler, D., et al., *Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation*. Mol Cell Biol, 2000. **20**(24): p. 9103-12.
39. Girard, A. and G.J. Hannon, *Conserved themes in small-RNA-mediated transposon control*. Trends Cell Biol, 2008. **18**(3): p. 136-48.
40. Watanabe, T., et al., *Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes*. Genes Dev, 2006. **20**(13): p. 1732-43.
41. Aravin, A.A., G.J. Hannon, and J. Brennecke, *The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race*. Science, 2007. **318**(5851): p. 761-4.

42. Yang, A.S., et al., *A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements*. Nucleic Acids Res, 2004. **32**(3): p. e38.
43. Schulz, W.A., C. Steinhoff, and A.R. Florl, *Methylation of endogenous human retroelements in health and disease*. Curr Top Microbiol Immunol, 2006. **310**: p. 211-50.
44. Whitelaw, E. and D.I. Martin, *Retrotransposons as epigenetic mediators of phenotypic variation in mammals*. Nat Genet, 2001. **27**(4): p. 361-5.
45. Han, J.S., S.T. Szak, and J.D. Boeke, *Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes*. Nature, 2004. **429**(6989): p. 268-74.
46. Yoder, J.A., C.P. Walsh, and T.H. Bestor, *Cytosine methylation and the ecology of intragenomic parasites*. Trends Genet, 1997. **13**(8): p. 335-40.
47. Lee, E., et al., *Landscape of Somatic Retrotransposition in Human Cancers*. Science, 2012.
48. Baillie, J.K., et al., *Somatic retrotransposition alters the genetic landscape of the human brain*. Nature, 2011. **479**(7374): p. 534-7.
49. Miki, Y., et al., *Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer*. Cancer Res, 1992. **52**(3): p. 643-5.
50. van den Hurk, J.A., et al., *L1 retrotransposition can occur early in human embryonic development*. Hum Mol Genet, 2007. **16**(13): p. 1587-92.
51. Hancks, D.C. and H.H. Kazazian, Jr., *Active human retrotransposons: variation and disease*. Curr Opin Genet Dev, 2012. **22**(3): p. 191-203.
52. Mills, R.E., et al., *Which transposable elements are active in the human genome?* Trends Genet, 2007. **23**(4): p. 183-91.
53. Bennett, E.A., et al., *Active Alu retrotransposons in the human genome*. Genome Res, 2008. **18**(12): p. 1875-83.
54. Wei, W., et al., *Human L1 retrotransposition: cis preference versus trans complementation*. Mol Cell Biol, 2001. **21**(4): p. 1429-39.
55. Esnault, C., J. Maestre, and T. Heidmann, *Human LINE retrotransposons generate processed pseudogenes*. Nat Genet, 2000. **24**(4): p. 363-7.
56. Nelson, H.H., C.J. Marsit, and K.T. Kelsey, *Global methylation in exposure biology and translational medical science*. Environ Health Perspect, 2011. **119**(11): p. 1528-33.
57. Schmid, C.W., *Does SINE evolution preclude Alu function?* Nucleic Acids Res, 1998. **26**(20): p. 4541-50.
58. Nellaker, C., et al., *Transactivation of elements in the human endogenous retrovirus W family by viral infection*. Retrovirology, 2006. **3**: p. 44.
59. Laska, M.J., et al., *Expression of HERV-Fc1, a human endogenous retrovirus, is increased in patients with active multiple sclerosis*. J Virol, 2012. **86**(7): p. 3713-22.
60. Baccarelli, A., et al., *Ischemic heart disease and stroke in relation to blood DNA methylation*. Epidemiology, 2010. **21**(6): p. 819-28.
61. Kim, K.Y., et al., *Association of low-dose exposure to persistent organic pollutants with global DNA hypomethylation in healthy Koreans*. Environ Health Perspect, 2010. **118**(3): p. 370-4.
62. Smolarek, I., et al., *Global DNA methylation changes in blood of patients with essential hypertension*. Med Sci Monit, 2010. **16**(3): p. CR149-155.
63. Kitkumthorn, N., et al., *LINE-1 methylation in the peripheral blood mononuclear cells of cancer patients*. Clin Chim Acta, 2012. **413**(9-10): p. 869-74.



64. Baccarelli, A., et al., *Rapid DNA methylation changes after exposure to traffic particles*. Am J Respir Crit Care Med, 2009. **179**(7): p. 572-8.
65. Bollati, V., et al., *Changes in DNA methylation patterns in subjects exposed to low-dose benzene*. Cancer Res, 2007. **67**(3): p. 876-80.
66. Tarantini, L., et al., *Effects of particulate matter on genomic DNA methylation content and iNOS promoter methylation*. Environ Health Perspect, 2009. **117**(2): p. 217-22.
67. Rossella, F., et al., *Urinary polycyclic aromatic hydrocarbons and monohydroxy metabolites as biomarkers of exposure in coke oven workers*. Occup Environ Med, 2009. **66**(8): p. 509-16.
68. Valinluck, V., et al., *Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2)*. Nucleic Acids Res, 2004. **32**(14): p. 4100-8.
69. Baccarelli, A., et al., *Repetitive element DNA methylation and circulating endothelial and inflammation markers in the VA normative aging study*. Epigenetics, 2010. **5**(3).
70. Pilsner, J.R., et al., *Associations of plasma selenium with arsenic and genomic methylation of leukocyte DNA in Bangladesh*. Environ Health Perspect, 2011. **119**(1): p. 113-8.
71. Wilhelm, C.S., et al., *Implications of LINE1 methylation for bladder cancer risk in women*. Clin Cancer Res, 2010. **16**(5): p. 1682-9.
72. Wright, R.O., et al., *Biomarkers of lead exposure and DNA methylation within retrotransposons*. Environ Health Perspect, 2010. **118**(6): p. 790-5.
73. Rusiecki, J.A., et al., *Global DNA hypomethylation is associated with high serum-persistent organic pollutants in Greenlandic Inuit*. Environ Health Perspect, 2008. **116**(11): p. 1547-52.
74. Li, E., *Chromatin modification and epigenetic reprogramming in mammalian development*. Nat Rev Genet, 2002. **3**(9): p. 662-73.
75. Sunder Raman, R., P.K. Hopke, and T.M. Holsen, *Characterization of fine aerosol and its inorganic components at two rural locations in New York State*. Environ Monit Assess, 2008. **144**(1-3): p. 351-66.
76. McVeigh, G.E., et al., *Nitric oxide modulation of blood vessel tone identified by arterial waveform analysis*. Clin Sci (Lond), 2001. **100**(4): p. 387-93.
77. Dolinoy, D.C., J.R. Weidman, and R.L. Jirtle, *Epigenetic gene regulation: linking early developmental environment to adult disease*. Reprod Toxicol, 2007. **23**(3): p. 297-307.
78. Fowler, B.A., et al., *Oxidative stress induced by lead, cadmium and arsenic mixtures: 30-day, 90-day, and 180-day drinking water studies in rats: an overview*. Biometals, 2004. **17**(5): p. 567-8.
79. Zhao, C.Q., et al., *Association of arsenic-induced malignant transformation with DNA hypomethylation and aberrant gene expression*. Proc Natl Acad Sci U S A, 1997. **94**(20): p. 10907-12.
80. Reichard, J.F., M. Schnekenburger, and A. Puga, *Long term low-dose arsenic exposure induces loss of DNA methylation*. Biochem Biophys Res Commun, 2007. **352**(1): p. 188-92.
81. Sciandrello, G., et al., *Arsenic-induced DNA hypomethylation affects chromosomal instability in mammalian cells*. Carcinogenesis, 2004. **25**(3): p. 413-7.

82. Chen, H., et al., *Chronic inorganic arsenic exposure induces hepatic global and individual gene hypomethylation: implications for arsenic hepatocarcinogenesis*. Carcinogenesis, 2004. **25**(9): p. 1779-86.
83. Pilsner, J.R., et al., *Genomic methylation of peripheral blood leukocyte DNA: influences of arsenic and folate in Bangladeshi adults*. Am J Clin Nutr, 2007. **86**(4): p. 1179-86.
84. Kinney, P.L., et al., *Airborne concentrations of PM(2.5) and diesel exhaust particles on Harlem sidewalks: a community-based pilot study*. Environ Health Perspect, 2000. **108**(3): p. 213-8.
85. Dioni, L., et al., *Effects of short-term exposure to inhalable particulate matter on telomere length, telomerase expression, and telomerase methylation in steel workers*. Environ Health Perspect, 2011. **119**(5): p. 622-7.
86. Baccarelli, A., et al., *Effects of particulate air pollution on blood pressure in a highly exposed population in Beijing, China: a repeated-measure study*. Environ Health, 2011. **10**: p. 108.
87. Rijnders, E., et al., *Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density*. Environ Health Perspect, 2001. **109 Suppl 3**: p. 411-7.
88. Lee, J., et al., *Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons*. Gene, 2007. **390**(1-2): p. 18-27.
89. Xing, J., et al., *Alu element mutation spectra: molecular clocks and the effect of DNA methylation*. J Mol Biol, 2004. **344**(3): p. 675-82.
90. van de Lagemaat, L.N., P. Medstrand, and D.L. Mager, *Multiple effects govern endogenous retrovirus survival patterns in human gene introns*. Genome Biol, 2006. **7**(9): p. R86.
91. Batzer, M.A. and P.L. Deininger, *Alu repeats and human genomic diversity*. Nat Rev Genet, 2002. **3**(5): p. 370-9.
92. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes*. J Mol Biol, 1987. **196**(2): p. 261-82.
93. Baccarelli, A., et al., *[DNA methylation analysis in environmental and occupational cancer research]*. G Ital Med Lav Ergon, 2005. **27**(3): p. 267-71.
94. Choi, S.H., et al., *Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer*. Int J Cancer, 2009. **125**(3): p. 723-9.
95. Bustin, S.A., *Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays*. J Mol Endocrinol, 2000. **25**(2): p. 169-93.
96. Dussault, A.A. and M. Pouliot, *Rapid and simple comparison of messenger RNA levels using real-time PCR*. Biol Proced Online, 2006. **8**: p. 1-10.
97. Vitzthum, F., et al., *A quantitative fluorescence-based microplate assay for the determination of double-stranded DNA using SYBR Green I and a standard ultraviolet transilluminator gel imaging system*. Anal Biochem, 1999. **276**(1): p. 59-64.
98. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method*. Methods, 2001. **25**(4): p. 402-8.
99. Medstrand, P., L.N. van de Lagemaat, and D.L. Mager, *Retroelement distributions in the human genome: variations associated with age and proximity to genes*. Genome Res, 2002. **12**(10): p. 1483-95.

100. Gu, Z., et al., *Densities, length proportions, and other distributional features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence*. Gene, 2000. **259**(1-2): p. 81-8.
101. Choi, S., et al., *Identification of toxicological biomarkers of di(2-ethylhexyl) phthalate in proteins secreted by HepG2 cells using proteomic analysis*. Proteomics, 2010. **10**(9): p. 1831-46.
102. Yang, N. and H.H. Kazazian, Jr., *L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells*. Nat Struct Mol Biol, 2006. **13**(9): p. 763-71.